

Studies of UHRF1-mediated mechanisms regulating DNA methyltransferase 1



Martha Smets
München 2017

Erstgutachter: Prof. Dr. Heinrich Leonhardt
Zweitgutachter: Prof. Dr. Bettina Kempkes

Tag der Abgabe: 06.06.2017

Tag der mündlichen Prüfung: 06.12.2017

Studies of UHRF1-mediated mechanisms regulating DNA methyltransferase 1



Dissertation
der Fakultät für Biologie
der Ludwig-Maximilians-Universität München

vorgelegt von
Martha Smets

München, den 06.06.2017

Table of Contents

Summary	1
Zusammenfassung	3
1 Introduction	5
1.1 Epigenetic regulation	5
1.2 Chromatin structure and histone modifications	6
1.2.1 Histone modifications associated with transcriptional activity	8
1.2.2 Histone modifications associated with transcriptional repression.....	9
1.3 DNA methylation	10
1.3.1 Setting DNA methylation.....	10
1.3.2 DNA modification dynamics during development.....	17
1.4 DNA modifications in neurogenesis and neurodegeneration.....	21
1.4.1 DNA modification dynamics during embryonic and adult neurogenesis..	21
1.4.2 DNA methylation and neurodegenerative diseases	23
1.5 New advances in genome engineering techniques	25
1.6 New advances in transcriptome analysis	27
1.7 Aims of this work	28
2 Results	31
2.1 DNMT1 mutations found in HSN1E patients affect interaction with UHRF1 and neuronal differentiation	31
2.2 DNA methylation requires a DNMT1 ubiquitin interacting motif (UIM) and histone ubiquitination	57
2.3 A modular open platform for systematic functional studies under physiological conditions.....	93
2.4 Comparative analysis of single-cell RNA sequencing methods.....	129
2.5 Ubiquitome analysis reveals PAF15 as a specific ubiquitination target of UHRF1 in embryonic stem cells.....	165
3 Discussion	193
3.1 DNMT1 mutations found in HSN1E patients affect interaction with UHRF1 and neuronal differentiation	193
3.2 DNA methylation requires a DNMT1 ubiquitin interacting motif (UIM) and histone ubiquitination	198
3.3 A modular open platform for systematic functional studies under physiological conditions.....	202
3.4 Comparative Analysis of Single-Cell RNA Sequencing Methods	206
3.5 Ubiquitome analysis reveals PAF as a specific ubiquitination target of UHRF1 in embryonic stem cells	208
4 Annex	213
4.1 References	213
4.2 Abbreviations	231
4.3 Supplementary Material	236
4.4 Curriculum vitae	238
4.5 Contributions.....	241

Summary

Epigenetic processes include the establishment, maintenance and coordinated change of DNA and histone modifications to shape chromatin structure across cell divisions without affecting the DNA sequence. This thesis addresses the regulation of DNA methylation with particular focus on the interplay between epigenetic factors DNMT1 and UHRF1, their role during development and disease as well as technology advancements in genome editing and transcriptomics towards better understanding of chromatin biology.

The field of epigenetics increasingly relies on genome editing and genome-wide transcriptomic analyses as key technologies to study and understand the regulation of individual genes, their protein products as well as systemic effects on gene expression. During my PhD, I co-established a multifunctional integrase (MIN) tag for rapid and versatile genome engineering. Based on CRISPR/Cas mediated manipulation, this approach enables efficient generation of multiple isogenic cell lines to study gene function under physiological conditions. Moreover, I participated in a comprehensive study comparing different library preparation methods for single-cell transcriptomic analyses.

The main study of this thesis focuses on mutations in the TS domain of DNMT1 that are linked to a neurodegenerative disease called HSN1E. Here, we characterized two disease related mutations with functional complementation assays in mouse embryonic stem cells. We showed that these mutations impair both interaction with UHRF1 and DNMT1 heterochromatin association resulting in decreased methylation levels. The next objective of this work was to investigate UHRF1-mediated modifications and their functional interplay with DNMT1 regulating DNA methylation. We identified H3K18 as a novel ubiquitination target of UHRF1. H3K18ub is essential for DNMT1 targeting and function. Consequently, we identified a ubiquitin interacting motif (UIM) within the TS domain of DNMT1, responsible for binding H3K18ub.

Finally, with *UHRF1* and *UHRF2* single knockout ESCs, we assayed for differential ubiquitination using mass spectrometry and identified novel non-histone ubiquitination targets. We show that UHRF1 ubiquitinates PAF15 at Lys 15 and Lys 24 and thereby promotes its binding to PCNA during late S phase.

Together my findings indicate a novel functional aspect of UHRF1 in regulating DNA methylation via H3K18ub and suggest a non-epigenetic role in DNA damage response. Despite sequence and domain similarity of UHRF1 and UHRF2, my results suggest that UHRF1/2 are involved in different cellular pathways, arguing for non-redundant functions.

Zusammenfassung

Epigenetische Prozesse umfassen die Etablierung, Aufrechterhaltung und koordinierte Veränderung von DNA- und Histon-Modifikationen. Desweiteren beeinflussen sie die Chromatinstruktur während des Zellzyklus, ohne dass die DNS-Sequenz verändert wird. Diese Doktorarbeit beschäftigt sich mit den molekularen Mechanismen der DNS-Methylierung mit besonderem Fokus auf dem funktionellen Zusammenspiel der epigenetischen Faktoren DNMT1 und UHRF1 und ihrer Funktion in der embryonalen Entwicklung. Desweiteren werden technologische Fortschritte in der Genommanipulation und Transkriptomanalyse zum besseren Verständnis der Chromatinbiologie beschrieben.

Das Feld der Epigenetik setzt zunehmend auf Genommanipulation und genomweite Transkriptomanalysen als Schlüsseltechnologien, um die Regulation einzelner Gene, ihre Proteinprodukte sowie systemische Effekte auf das zelluläre Transkriptom zu untersuchen und zu verstehen. Während meiner Promotion habe ich das *multifunctional integrase (MIN) tag* System für eine schnelle und vielseitige Genommanipulation mitetabliert. Basierend auf CRISPR/Cas-vermittelten Manipulationen ermöglicht dieser Ansatz eine effiziente Generierung von isogenen Zelllinien, um individuelle Gene unter physiologischen Bedingungen zu untersuchen. Darüber hinaus habe ich an einer umfassenden Studie teilgenommen, in deren Rahmen verschiedene Verfahren der Einzelzell-Transkriptomanalyse verglichen wurden.

Die Hauptstudie dieser Arbeit konzentriert sich auf Mutationen in der TS-Domäne von DNMT1, die mit einer neurodegenerativen Erkrankung namens HSN1E assoziiert sind. Hier haben wir zwei krankheitsbezogene Mutationen mit Komplementierungsexperimenten in embryonalen Stammzellen charakterisiert. Wir haben gezeigt, dass diese Mutationen sowohl die Bindung von DNMT1 an Heterochromatin als auch die Interaktion mit UHRF1 beeinträchtigen, was zu einem verringerten Methylierungslevel führt.

Das nächste Ziel dieser Arbeit war es, UHRF1-vermittelte posttranslationale Modifikationen und deren Einfluss auf die Regulation von DNMT1 zu untersuchen. Wir identifizierten H3K18 als eine neue Histonubiquitinierungsstelle, die von UHRF1 ubiquitiniert wird. H3K18ub ist essentiell für die Funktion von DNMT1. In der Folge haben wir ein *ubiquitin interacting motif* (UIM) innerhalb der TS-Domäne von DNMT1 identifiziert, das für die Bindung von H3K18ub verantwortlich ist.

Schließlich wurden embryonale UHRF1- und UHRF2-*knockout* Stammzellen auf differenzielle Ubiquitinierung mittels Massenspektrometrie untersucht und neue

Zusammenfassung

Ubiquitinierungssubstrate identifiziert. Wir zeigen, dass UHRF1 PAF15 ubiquitiniert und damit die Bindung von PAF15 an PCNA in der späten S-Phase fördert.

Die Ergebnisse dieser Arbeit zeigen einen neuen funktionellen Aspekt von UHRF1 bei der Regulation der DNS-Methylierung über H3K18ub und deuten auf eine Funktion bei der Reparatur von DNS-Schäden hin. Trotz der Ähnlichkeit von Sequenz und Domänenstruktur von UHRF1 und UHRF2 deuten die Ergebnisse darauf hin, dass UHRF1/2 an verschiedenen zellulären Prozessen beteiligt sind und unterschiedliche Funktionen haben.

1 Introduction

1.1 Epigenetic regulation

Although almost all cells of a multicellular organism are genetically homogeneous, they are structurally and functionally heterogeneous and give rise to multiple tissues (Bird 2002). To explain how different cellular morphology and function can arise from the totipotent zygote during development, the epigenome provides a superordinate layer of information (Waddington 1957; Jaenisch and Bird 2003). Central to the definition of epigenetics is the knowledge that genes carry regulatory information beyond their nucleotide sequences. Epigenetic mechanisms are either dynamic during early embryonic development or relatively stable, capable of being passed onto daughter cells (Smith and Meissner 2013). In higher order chromatin, epigenetic information is mediated by DNA and histone modifications and further histone variants and nucleosome positioning as well as RNA-mediated gene silencing (Figure 1). These mechanisms function in a highly complex regulatory system and deregulation leads to genomic instability and promotes tumorigenesis (Goldberg, Allis, and Bernstein 2007; Choi and Jong-Soo 2013). It has become clear that significant crosstalk exists between epigenetic modifications and pathways, which are connected in various combinations with each other that manifolds the complexity of epigenetic regulation (Reik 2007; Ponting, Oliver, and Reik 2009; Bannister and Kouzarides 2011). As DNA and histone modifications are the most prominent and well-studied epigenetic marks, they will be introduced in the next chapters.

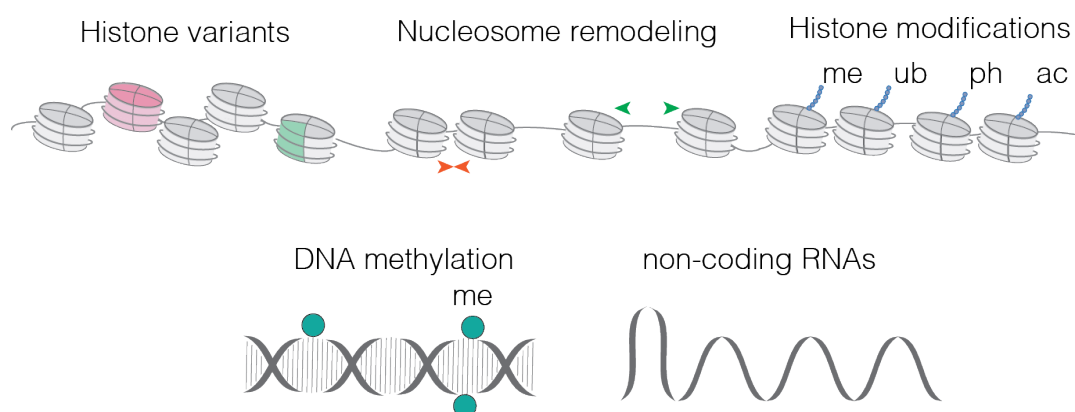


Figure 1. Epigenetic mechanisms. Gene expression is regulated by the crosstalk between DNA methylation and posttranslational histone modifications. The replacement of canonical histone proteins by histone variants and nucleosome remodeling can additionally alter the accessibility of chromatin. Further, non-coding RNAs contribute to epigenetic regulation.

1.2 Chromatin structure and histone modifications

The basic level of chromatin compaction in eukaryotic cells consist of DNA wrapped around a histone octamer forming nucleosomes that are folded into higher-order chromatin fibres (Luger and Richmond 1998; Kornberg and Lorch 1999). Two copies of histones H2A, H2B, H3, and H4 pairwise interact to form the barrel-shaped histone particle called histone octamer (Woodcock and Ghosh 2010). To further compact chromatin, a linker histone called H1 localizes near the DNA entry-exit sites of the core particle (Figure 2) (Thoma, Koller, and Klug 1979). Histones possess a globular structure, but also harbor a 20-35 amino acid long N-terminal unstructured peptide that protrudes from the surface of the nucleosome called histone tail (Luger et al. 1997). All histones can be posttranslationally modified, and the sites of modification are mainly on the histone tails.

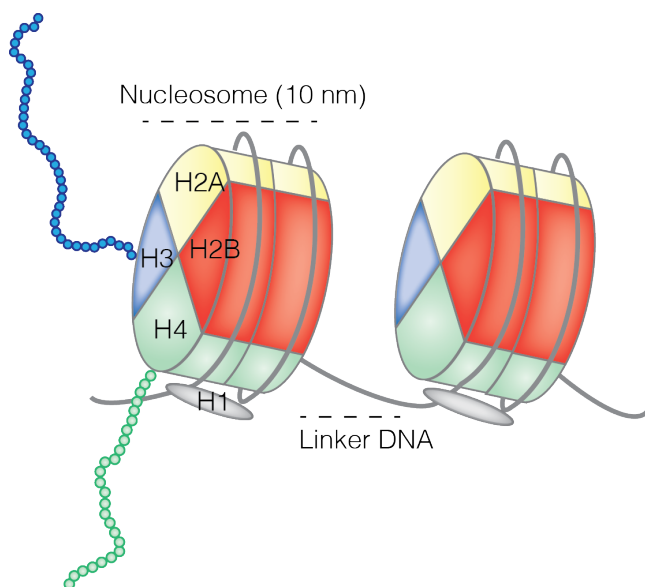


Figure 2. The nucleosome. 147 bp of DNA is wrapped around the histone octamer that consists of two histone H3, H4 and two heterodimers H2A and H2B. The linker histone H1 is stabilizing the nucleosome and is further compacting the chromatin. Shown are also histone tails protruding from H3 and H4.

So called 'writers' set modification marks such as acetylation, phosphorylation, ubiquitination, sumoylation, ADP ribosylation, deimination, proline isomerization as well as different degrees of methylation (Kouzarides 2007). Additionally, unmethylated, monomethylated but not di-methylated arginines can be converted to citrulline by deimination (Cuthbert et al. 2004).

The most abundant posttranslational modifications (PTMs) are set by different groups of writers including kinases, histone acetyltransferases (HATs) and histone methyltransferases (HMTs), which play important roles in development. Mutation or misregulation of writers is associated with genetic disorders as well as various cancers (T. Zhang, Cooper, and Brockdorff 2015).

Histones carrying certain PTMs are recognized by 'reader' modules that bind their target (Kouzarides 2007; Taverna et al. 2007). Methylated, acetylated, phosphorylated histones are recognized by chromo-like domains of the Royal family, bromodomains and a domain within 14-3-3 proteins, respectively. By reading the histone PTMs, either the binding protein itself or in cooperation with associated complexes initiate functional pathways. Histone PTMs can be stable or dynamic as they can be removed by the catalytic activity of erasers like histone deacetylases (HDACs), deubiquitinases (DUBs) and demethylases (Kouzarides 2007).

Histone PTMs influence gene expression by modulating chromatin structure. There are two suggested models for the mode-of-action of histone PTMs. First, by modifying the electrostatic charge of the histone, PTMs may induce a change in chromatin organization and state of condensation and thus regulate the accessibility for non-histone proteins to chromatin. One such modification is histone acetylation, which was the first PTM shown to correlate with gene expression (Allfrey, Faulkner, and Mirsky 1964). Acetylation neutralizes the positively charged lysine (K) residues of the histone N-terminus, weakening histone-DNA and internucleosome contacts and thereby reducing chromatin compaction. Together, this causes the nucleosomes to unfold and thereby enables the binding of the transcription machinery (Workman and Kingston 1998; Zentner and Henikoff 2013). Second, histone modifications influence gene expression by recruitment of regulatory factors that bind histone PTMs directly (Strahl and Allis 2000). For example, the bromodomains of several proteins involved in transcriptional activation bind to acetylated lysines of histone H3 and H4 (Jacobson et al. 2000). Histone modifications can also lead to the recruitment of DNA methyltransferases and thereby lead to DNA methylation and transcriptional repression (Tachibana et al. 2008). Trimethylated histone H3 on K9 (H3K9me3), a predominant mark for heterochromatin (Lachner et al. 2001), can be bound by the chromodomain of heterochromatin protein 1 (HP1), which can lead to repression of transcription (Danzer and Wallrath 2004).

Moreover, histone marks can regulate each other, giving rise to a dynamic epigenetic crosstalk. The presence of one modification can promote or inhibit the occurrence of one or more subsequent modifications on the same histone molecule, or between different histone molecules and across nucleosomes. Depending on the combination of histone PTMs, the functional output leading to activation or repression of the underlying gene can differ (Latham and Dent 2007; Strahl and Allis 2000). In the next chapter distinct patterns of histone marks associated with active and repressed transcriptional states will be introduced (Figure 3).

Introduction

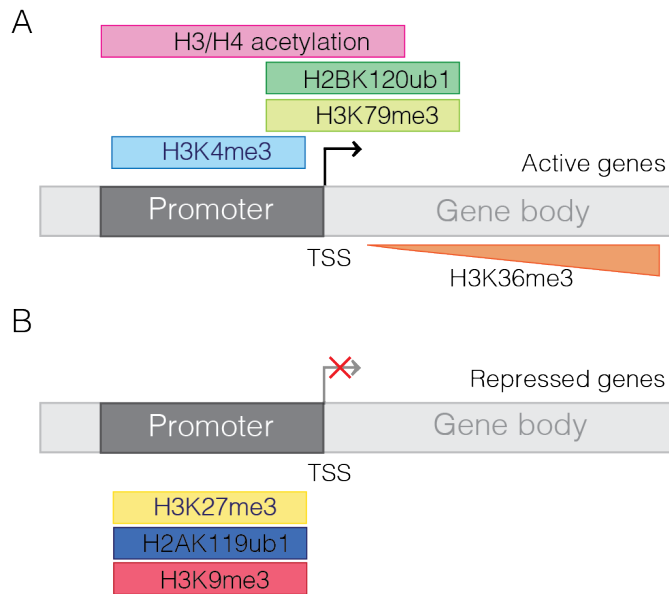


Figure 3. Histone posttranslational modifications of active and repressed genes. (A) PTMs found in promoters, transcriptional start sites and gene bodies of active genes are shown. (B) PTMs found in promoters of repressed genes are shown. Modified from (T. Zhang, Cooper, and Brockdorff 2015).

1.2.1 Histone modifications associated with transcriptional activity

Actively transcribed genes are usually found in euchromatic regions and are characterized by an array or combination of PTMs on histone tails (Figure 3A, B). Active enhancers are marked by H3K4me1 and H3K27ac and promoters as well as transcriptional start sites (TSS) of active genes are marked by a high density of H3K4me3 as well as H3 and H4 acetylation (Deckert and Struhl 2001; Liang et al. 2004; Barrera et al. 2008; Creighton et al. 2010). The distribution of H3K4me3 is highly coupled to regions of CpG- and GC-dense DNA, so called CpG islands (CGI) that are found in 50–70% of all promoters (Deaton and Bird 2011). All CGI promoters are marked with H3K4me3, and high levels of H3K4me3 correlated with high gene expression (Barski et al. 2007; Guenther et al. 2007). The gene bodies of euchromatic genes are enriched in H3 and H4 acetylation (Myers et al. 2001), H3K79me3 (Ng et al. 2003), and H2BK120ub1 (Batta et al. 2011; Ng, Dole, and Struhl 2003), as well as increasing H3K36me3 levels towards the 3' end (Pokholok et al. 2005; Neri et al. 2017). Chromatin marks associated with active genes can positively reinforce each other through various positive feedback mechanisms. In mammals, one example is the co-accumulation of H3K4me3 and H2BK120ub1 on gene bodies, as the knockdown of the H2BK120ub1-specific E3-ligases RNF20/40 lead to global reduction of H3K4me3 (J. Kim, Hake, and Roeder 2005).

1.2.2 Histone modifications associated with transcriptional repression

Repressed genes are usually found in heterochromatic regions and are characterized by a combination of PTMs on histone tails, for example methylation of residues K27 and K9 of H3 and the ubiquitination of H2A on K119 (Figure 3). H3K27me3 and H2AK119ub1 are associated with the formation of facultative heterochromatin, which is competent to interconvert between active and inactive states depending on the spatiotemporal context. H3K9me2/3 has important roles in the formation of constitutive heterochromatin, which is largely transcriptionally inert except for early developmental and pathological situations (Smith and Meissner 2013; T. Zhang, Cooper, and Brockdorff 2015). The methylation of H3K27 is set by Polycomb Repressive Complex 2 (PRC2), which consists of five subunits, EZH2 (enhancer of zeste homolog 2), EED (Embryonic Ectoderm Development), SUZ12 (suppressor of zeste 12 protein homolog), and the histone binding proteins RBBP4 and RBBP7 (Kuzmichev et al. 2002). Different states of methylation of H3K27, mono-, di-, and trimethylated, can all be set by PRC2 and have different biological functions. H3K27me1/2 is enriched on gene bodies and is associated with gene activation whereas H3K27me3 is associated with gene repression (Morey and Helin 2010; Ferrari et al. 2014). The monoubiquitination of H2AK119ub1 is set by RING1B, a subunit of Polycomb Repressive Complex 1 (PRC1) (Bhaumik, Smith, and Shilatifard 2007). Interestingly, both PRC1 and PRC2 together with their associated chromatin modification co-localize at many genomic regions, like promoters of developmentally regulated genes (Ku et al. 2008). The exact mode-of-action of PRC1 and PRC2 and their hierarchical recruitment model is still under investigation (Tavares et al. 2012; Schoeftner et al. 2006; L. Wang et al. 2004; Ren, Vincenz, and Kerppola 2008; Blackledge et al. 2014; Cooper et al. 2014; Kalb et al. 2014).

Interestingly, CGI promoters linked to developmental and lineage-specific genes in embryonic stem cells (ESCs), harbor the repressive H3K27me3 as well as the active H3K4me3 mark. This PTMs combination is called bivalency and it is suggested that it renders genes poised for immediate activation upon differentiation (Bernstein et al. 2006; Azuara et al. 2006).

1.3 DNA methylation

DNA methylation is among the best-studied epigenetic marks and was the first epigenetic mark found to correlate with gene expression. Methylated mammalian promoters are thought to lead to transcriptional repression either by inhibiting transcription factors to bind or by further compacting the chromatin. DNA methylation is highly conserved among different species such as most plants, animals as well as fungi and is essential for mammalian development (Smith and Meissner 2013). In mammals, DNA methylation occurs predominantly at the C5 position of cytosine in a symmetrical CpG dinucleotide context resulting in 5-methylcytosine (5mC), however, low non-CpG methylation levels in a CpA and CpT context were described in ESCs and are also prevalent in oocytes and neurons (Patil, Ward, and Hesson 2014). In *Drosophila melanogaster*, DNA methylation was mainly found in a CpT and in plants additionally in a CpA and CpC context (Bird 2002; Law and Jacobsen 2010).

In the human genome, 60-80% of the roughly 28 million CpGs are generally methylated. Less than 10% of CpGs occur in CGIs, are usually hypomethylated and associated with TSSs of housekeeping and development associated genes (Smith and Meissner 2013). A small subset of the CGIs gains methylation during development leading to a stable transcriptional repression.

However, DNA methylation is also found in transcriptionally active gene bodies and at exon-intron boundaries, where it was suggested to regulate co-transcriptional splicing and to inhibit spurious transcription initiation (Wolf et al. 1984; Laurent et al. 2010; Maunakea et al. 2013; Neri et al. 2017).

1.3.1 Setting DNA methylation

The enzymes responsible for the addition of the methyl group to cytosines belong to the protein family of DNA methyltransferases (DNMTs). In vertebrates, different DNMT members have been described, which harbor a catalytically active C-terminal domain. All of them contain a highly conserved catalytic domain, which includes 10 sequence motives that are also found in prokaryotic cytosine-C5 DNA methyltransferases (Figure 4A) (Goll and Bestor 2005; Qin, Leonhardt, and Pichler 2011).

Reaction mechanism

All catalytically active DNMTs share a common reaction mechanism, requiring S-adenosyl-L-methionine (SAM) as a methyl-donor. In the first step of the methylation reaction, DNMTs bind to the DNA and flip out the target base. Subsequently, they form a covalent complex by a conserved cysteine nucleophile with the C6 position of cytosine. This results in the activation of the C5 atom and the methyl group is transferred from the cofactor SAM, which serves as the common cellular methyl group donor. In the last reaction step, the covalent bond is resolved by β -elimination and the enzyme released from the DNA (Figure 4B) (X. Cheng and Blumenthal 2008).

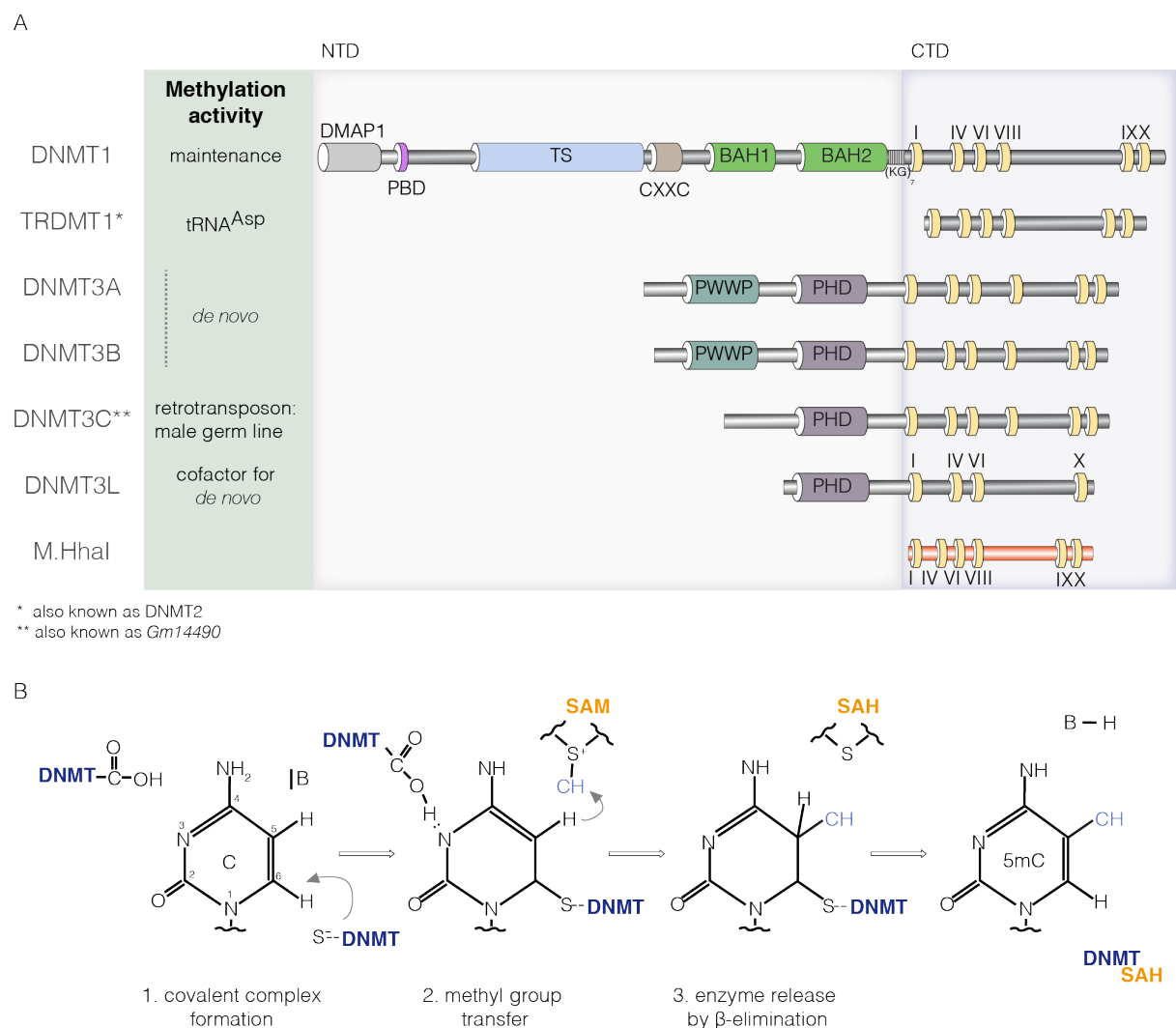


Figure 4. Domain structure of DNMTs and the methylation reaction of the mammalian DNMT family. (A) Schematic outline of the domain architecture of mammalian DNMTs in comparison to the prokaryotic DNMT, M.HhaI. All DNMTs except DNMT3L harbor a catalytically active C-terminal domain. DMAP1: a DNA methyltransferase associated protein 1 (DMAP1)-binding domain; PBD: PCNA binding domain; TS: Targeting sequence; CXXC: zinc finger domain; BAH: Bromo-adjacent homology domain; PWWP: Pro-Trp-Trp-Pro motif containing

Introduction

domain; PHD: plant homeodomain. Modified from (Rottach, Leonhardt, and Spada 2009; Qin, Leonhardt, and Pichler 2011). **(B)** Methylation reaction of the DNMT family. After the formation of a covalent complex of the DNMTs with the C6 position of the cytosine (C), the methyl group is transferred from the cofactor SAM to the C5 position of the cytosine, leaving the demethylated cofactor S-adenosyl-L-homocystein (SAH). DNMTs are released by β -elimination. Modified from (Goll and Bestor 2005).

The DNMT3 family and DNMT2

Upon differentiation, many pluripotency associated genes such as *Oct4* (also known as *Pou5f1*) have to be silenced by epigenetic marks to ensure correct differentiation (Feldman et al. 2006). DNA methylation represents the final step in gene repression after histone PTMs on already transiently silenced loci thereby finalizing the epigenetic status of silent chromatin (Bird 2002).

DNMT3A and DNMT3B show catalytic activity on unmethylated DNA and are responsible for establishing *de novo* methylation during embryogenesis and gametogenesis. In ESCs and germ cells, DNMT3A and DNMT3B are highly expressed, while in differentiated cells and tissues they are downregulated (Daisuke Watanabe et al. 2002). In *DNMT3A/DNMT3B* double-knockout ESCs, methylation of imprinted regions is not affected, while newly introduced retroviral elements remain hypomethylated (Okano et al. 1999; Kaneda et al. 2004).

In vivo studies of either *DNMT3A*- or *DNMT3B*- knockout mice demonstrate that DNMT3B has an essential role during early developmental stages, while DNMT3A is important for methylation in later development. *DNMT3A*^{-/-} mice die four weeks after birth and *DNMT3B*^{-/-} mice display several developmental defects and are not viable, indicating that DNMT3B has an essential function during early development, whereas DNMT3A is important for methylation in later developmental stages (Okano et al. 1999).

Mutations in the *DNMT3A* or *DNMT3B* genes are associated with human diseases. Patients with mutations in *DNMT3B* suffer from the rare immunodeficiency, centromere instability and facial abnormalities (ICF) syndrome and are characterized by DNA hypomethylation of CpG sites in pericentromeric satellite regions as well as on the inactive X chromosome (Hansen et al. 1999; Bourc'h et al. 1999). Mutations in the functional domains of human *DNMT3A* cause overgrowth syndromes, intellectual disabilities and facial dysmorphism. None of the overgrowth-associated mutations seem to affect DNA binding ability of DNMT3A. Instead, their positions within the DNMT3A domains suggest that recognition of unmethylated H3K4, intramolecular interactions within DNMT3A and histone binding is disturbed and thereby *de novo* methylation is impaired (Tatton-Brown et al. 2014).

Both DNMT3A and DNMT3B have a long regulatory N-terminal domain (NTD) fused to a catalytic C-terminal domain (CTD). The NTD contains a Pro-Trp-Trp-Pro (PWWP) domain that enhances methyltransferase activity by binding H3K36me3 and thus is targeted to heterochromatin (Fuks et al. 2001; Ge et al. 2004; Dhayalan et al. 2010). Interaction with multiple chromatin-associated proteins, such as heterochromatin protein (HP1), histone deacetylase 1 (HDAC1) and histone methyltransferase SUV39H1 is mediated by the cysteine-rich plant homeodomain (PHD) (Fuks et al. 2001, 2003).

One member of the DNMT family, DNMT3L, lacks methylation activity but cooperates with DNMT3A and DNMT3B and stimulates their methylation activity (Chedin, Lieber, and Hsieh 2002; Gowher et al. 2005). DNMT3L regulates methylation at gene bodies of housekeeping genes positively and at bivalent promoters negatively (Neri et al. 2013) and acts downstream of the piRNA pathway (Aravin et al. 2008; Bourc'his and Bestor 2004). The inactivation of DNMT3L results in hypomethylation and reactivation of retrotransposons, meiotic failure and male sterility (Zamudio and Bourc'his 2010).

Recently, an additional *de novo* DNMT enzyme was discovered: DNMT3C, which evolved via a duplication event of DNMT3B in rodent genomes and is responsible for methylating promoters of evolutionarily young retrotransposons to ensure male fertility of mice (Barau et al. 2016). Remarkably, DNMT3C shares 70% identity with DNMT3B, while DNMT3A and DNMT3B are 46% identical.

TRDMT1 (also known as DNMT2) comprises only a catalytic domain, is involved in methylation of cytoplasmic tRNA^{Asp} and shows very weak DNA methyltransferase activity (Hermann, Schmitt, and Jeltsch 2003; Goll et al. 2006). However, TRDMT1 may be responsible for rare cytosine methylation at sequence contexts other than CpG (Kunert et al. 2003; Borsatti and Mandrioli 2004).

DNMT1

The maintenance of tissue-specific DNA methylation patterns to future cell generations is mediated by DNMT1 after each cycle of replication (Bird 2002; Smith and Meissner 2013). DNMT1 is essential for development as a catalytic *DNMT1* mutation in mouse embryos results in hypomethylation, delayed development and lethality during gestation (E. Li, Bestor, and Jaenisch 1992). Reduction of *DNMT1* expression in mice to 10% of wildtype (wt) level is compatible with viability but results in genome-wide hypomethylation and chromosomal instability leading to severe T-cell lymphomas (Gaudet et al. 2003). In contrast to the CTD of DNMT3A and DNMT3B, the CTD of DNMT1

Introduction

alone is not sufficient for methylation activity but requires a large part of the NTD (Margot et al. 2000; Anton and Bultmann 2017).

The very N-terminus of DNMT1 has a DNA methyltransferase associated protein 1 (DNMT1)-binding domain that binds HDAC2 and mediates the interaction of the transcriptional repressor DNMT1 with DNMT1 (Rountree, Bachman, and Baylin 2000). In late S phase, DNMT1 co-localizes and interacts with HDAC2 as well as HDAC1 coupling histone deacetylation activity with DNA methylation at replicating heterochromatin (Fuks et al. 2000; Rountree, Bachman, and Baylin 2000).

DNMT1 localization to sites of DNA replication is mediated by the PCNA binding domain (PBD) in early to mid S phase, while the targeting sequence (TS) domain mediates heterochromatin binding during late S and G2 (Leonhardt et al. 1992; Chuang et al. 1997; Easwaran et al. 2004). In addition, DNMT1 harbors a CXXC type zinc finger domain (CXXC) that binds unmethylated CpG sites (Fatemi et al. 2001; Pradhan et al. 2008; Frauer et al. 2011). Interestingly, structural insights find an inhibitory role of the CXXC-bromo adjacent homology domain 1 (BAH1) linker, which inhibits binding and *de novo* methylation of unmethylated CpG sites during maintenance methylation. This inhibitory mechanism is mediated by the complex of unmodified DNA with the CXXC-BAH1 linker, which blocks the access of CTD to the target CpG site (J. Song et al. 2011).

The two BAH domains are involved in several protein-protein interactions such as with HP1 β and ubiquitin-specific processing protease 7 (USP7) (Fuks et al. 2003; Du et al. 2010; Qin, Leonhardt, and Spada 2011).

The NTD and the CTD are fused by a linker that contains several lysyl-glycyl dipeptide repeats ((KG)₇) providing a flexible region between the two domains of DNMT1. The CTD of DNMT1 contains all ten conserved motifs for the methyl group transfer but the intramolecular interactions with the NTD are essential for allosteric activation and methylation activity (Margot et al. 2000; Fatemi et al. 2001; Takeshita et al. 2011).

DNMT1 is not only subjected to protein interactions but also to multiple PTMs, which contribute to proper regulation of DNA methylation and DNMT1 stability (Qin, Leonhardt, and Pichler 2011). Throughout the cell cycle, the stability, abundance and catalytic activity *in vivo* of DNMT1 is controlled by ubiquitination and acetylation as well as by phosphorylation with subsequent methylation (B. Lee and Muller 2009; Du et al. 2010; Estève et al. 2011; Felle et al. 2011; Qin, Leonhardt, and Pichler 2011). Long non-coding RNAs have also been shown to regulate DNMT1 activity (Di Ruscio et al. 2013).

Taken together, DNMT1 is regulated in a highly complex manner by PTMs and intramolecular protein interactions. Besides intramolecular protein-protein binding, also intermolecular protein interactions serve as a prerequisite for DNMT1 activation, which have to be overcome by conformational changes before the methylation reaction can occur.

The interaction of DNMT1 with Ubiquitin-like, containing PHD and RING finger domains 1 (UHRF1, also known as 95 kDa mouse nuclear protein (Np95)) releases the TS domain and enables catalytic activity of the CTD (Bashtrykov, Jankevicius, et al. 2014). Therefore, UHRF1 has emerged as an essential cofactor for maintenance DNA methylation.

UHRF protein family

UHRF1 binds, flips out hemimethylated DNA and recruits DNMT1 to its substrate sites (Arita et al. 2008; Avvakumov et al. 2008; Bostick et al. 2007; Hashimoto et al. 2008). Consistently, UHRF1 deficient embryos phenocopy the DNA hypomethylation and early embryonic lethality of *DNMT1*- knockout embryos, indicating that the multi-domain protein UHRF1 (Figure 5) is a key factor in maintenance methylation (Sharif et al. 2007). The SRA domain of UHRF1 mediates the direct interaction with DNMT1 and is essential for allosteric activation of DNMT1, which enables binding of substrate DNA to the CTD (Syeda et al. 2011; Takeshita et al. 2011; Bashtrykov, Rajavelu, et al. 2014; Berkuyrek et al. 2014). In addition to hemimethylated DNA, UHRF1 binds to methylated H3K9 via its tandem Tudor domain (TTD) (Citterio et al. 2004; Karagianni et al. 2008; Rottach et al. 2010). Via its plant homeodomain (PHD), which was previously implicated in transcriptional regulation and heterochromatin organization, UHRF1 was shown to bind the unmodified N-terminus of histone H3 and via its SRA domain target DNMT1 to hemimethylated sites (Papait et al. 2007; Rajakumara et al. 2011).

Crystal structures of the linked TTD–PHD bound to H3K9me3 peptides revealed that UHRF1 can simultaneously bind the unmodified H3 N-terminus and H3K9me3 on a single histone H3 tail through this cooperative recognition module (J. Cheng et al. 2013). However, the accessibility of the different histone H3-binding domains is allosterically regulated by binding of phosphatidylinositol 5-phosphate (PI5P) to a polybasic site in a linker region of UHRF1 between the SRA and RING domain. PI5P controls access to the TTD and PHD domain and so the interaction with unmodified H3 or H3K9me3, respectively, which might influence UHRF1 heterochromatin localization and function (Gelato et al. 2014).

Introduction

In addition, UHRF1 harbors a RING domain that regulates DNMT stability together with USP7 and ubiquitinates histone tails (Citterio et al. 2004). However, the exact specificity of the RING domain is still under investigation.

Taken together, the cooperative binding of hemimethylated DNA and repressive histone marks by UHRF1 targets DNMT1 to newly synthesized DNA in heterochromatin after replication (X. Liu et al. 2013). Furthermore, UHRF1 was shown to interact with DNMT3A, DNMT3B and several histone-modifying enzymes like HDAC1 or the histone methyltransferase G9A (Achour et al. 2008; J. K. Kim et al. 2009; Meilinger et al. 2009).

Besides epigenetic regulation, UHRF1 is also involved in a variety of processes ranging from DNA damage response to DNA replication (Fujimori et al. 1998; Muto 2002; Uemura et al. 2000).

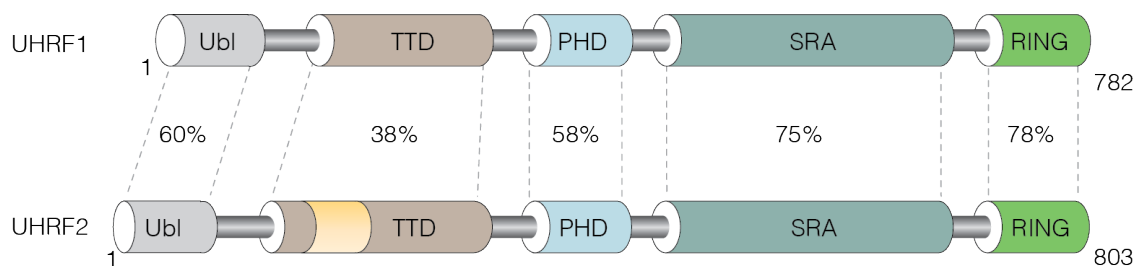


Figure 5. The UHRF family structure. The UHRF family harbors ubiquitin-like (Ubl) domain, a tandem Tudor domain (TTD), a plant homeodomain (PHD), a SET and RING-associated (SRA) domain and a really interesting new gene (RING) domain. Homology between UHRF1 and its paralogue UHRF2 are indicated. Numbers indicate the length of the murine proteins in amino acids.

The paralogue UHRF2 shares a high domain structure similarity with UHRF1 (Figure 5) (Bronner et al. 2007). Interestingly, both genes are expressed in opposite patterns. UHRF1 is highly expressed in undifferentiated cells, whereas UHRF2 is upregulated during differentiation and highly expressed in somatic cells especially neuronal cells. Although both UHRF1 and UHRF2 interact with DNMT1, UHRF2 cannot target DNMT1 to pericentric heterochromatin (PCH) in S phase and ectopic expression of UHRF2 in *UHRF1*^{-/-} ESCs did not restore DNA methylation levels arguing for non-redundant functions (Pichler et al. 2011; J. Zhang et al. 2011; R. Chen et al. 2017). Recently however, both UHRF1 and UHRF2 were shown to be negative regulators of DNA *de novo* methylation by DNMT3A by functioning as E3 ligases promoting DNMT3A ubiquitination and thereby subsequent degradation (Jia et al. 2016). Very recently, it was shown that *UHRF2* knockout mice are viable as well as fertile and exhibit no DNA

methylation defect. Except a partial reduction in spatial memory acquisition and retention, no gross developmental defect was observed (R. Chen et al. 2017).

In combination with binding to H3K9me3 mediated by the TTD, UHRF2 displays a preference for hemimethylated DNA. The localization and *in vivo* binding characteristics of UHRF2 were described to require an intact TTD and depend on H3K9me3 recognition but not on DNA methylation (Pichler et al. 2011). Interestingly, UHRF2 has 3-fold higher binding affinity to 5-hydroxymethylcytosine (5hmC) than 5mC and is a specific binder of 5hmC in neuronal progenitor cells (NPCs) (Spruijt et al. 2013; Zhou et al. 2014). Further, UHRF2 is recruited to 5hmC by the zinc finger protein 618 (ZNF618) indicating that ZNF618 might be important for the function of UHRF2 as a specific 5hmC reader *in vivo* (Y. Liu et al. 2016). UHRF2 interacts with DNMT3A and DNMT3B, HDAC1, G9a and H3K9me3 (Pichler et al. 2011; J. Zhang et al. 2011).

Besides epigenetic regulation, UHRF2 represents a nodal point in the cell cycle network as it interacts with cell cycle proteins including cyclins, cyclin-dependent kinases, retinoblastoma 1 protein (RB1), tumor protein p53, and PCNA (Mori et al. 2011, 2012). Taken together, UHRF2 might be a link between the epigenetic network and cell cycle regulation.

1.3.2 DNA modification dynamics during development

DNA methylation was thought to be a stable epigenetic mark due to its chemical nature, but dynamic changes of DNA methylation can be observed during early development (Figure 6). The maternal genome gradually loses DNA methylation during subsequent cell divisions as the oocyte-specific isoform of DNMT1 (DNMT1o) is actively excluded from the nucleus (passive dilution) (Cardoso and Leonhardt 1999). However, imprinted genes are excluded from passive dilution and are only demethylated in primordial germ cells (PGCs) (Monk, Boubelik, and Lehnert 1987; Howlett and Reik 1991). In contrast, active DNA demethylation occurs in sperm-derived paternal pronuclei genome-wide before the two pronuclei merge (Mayer et al. 2000). Imprinting control regions such as the H19 promoter, IAP retrotransposons and centromeric as well as pericentromeric regions avoid active DNA demethylation in the paternal genome (Olek and Walter 1997; Rougier et al. 1998; Lane et al. 2003).

Introduction

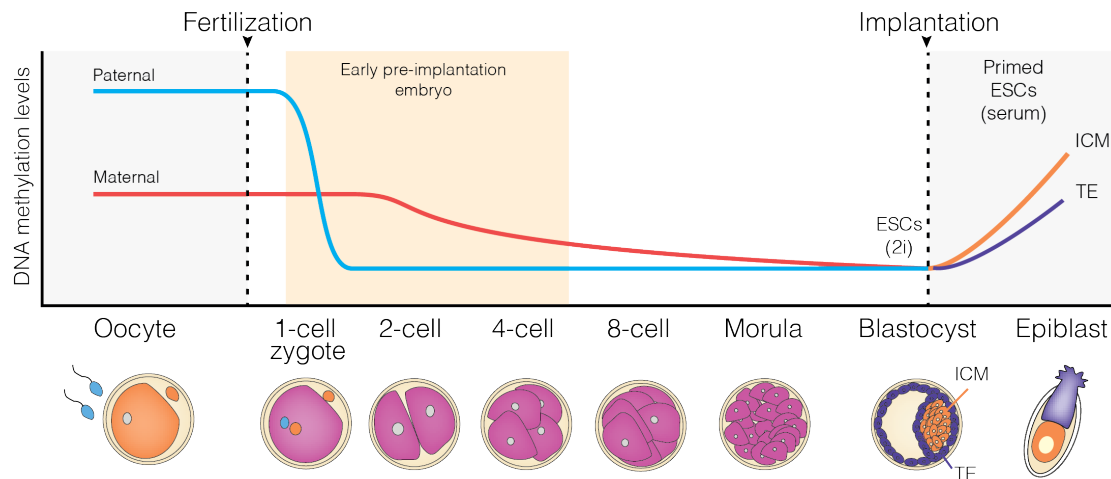


Figure 6. DNA modification dynamics during development. Shortly after fertilization, the paternal genome undergoes rapid genome-wide active demethylation, whereas the maternal genome stays at its methylated state. During the following first cell divisions, DNA methylation in the maternal genome is diluted by passive demethylation in the absence of DNMT1. Modified from (H. Wu and Zhang 2014).

Although, active DNA demethylation was already described 1982, the contributing enzymes have not been identified until recently (Gjerset and Martin 1982; Tahiliani et al. 2009). The responsible enzymes were first described in a different biological context: as a fusion partner of the MLL gene in acute myeloid leukemia (AML) in a t(10;11)q(22;23) translocation event (TET: ten-eleven-translocation) (Lorsbach et al. 2003). The family of TET proteins (TET1, TET2 and TET3) converts 5mC to 5hmC, 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC). The higher oxidized cytosine variants, 5fC and 5caC, can be excised by the Thymine-DNA glycosylase (TDG) (He et al. 2011; Maiti and Drohat 2011). This process is thought to be an intermediate step of active DNA demethylation (Tahiliani et al. 2009; Kriaucionis and Heintz 2009). Loss of genome-wide paternal DNA methylation concurs with a rapid increase in 5hmC, 5fC and 5caC, suggesting that TET-mediated 5mC oxidation contributes to active demethylation (Inoue and Zhang 2011; Iqbal et al. 2011).

The epigenetic mechanisms in early development are of great interest to the scientific field and are still under investigation and great debate. The accumulation of 5hmC on the paternal DNA has been mainly attributed to the activity of TET3, the only TET member present at this stage (Gu et al. 2011; Iqbal et al. 2011; Wossidlo et al. 2011).

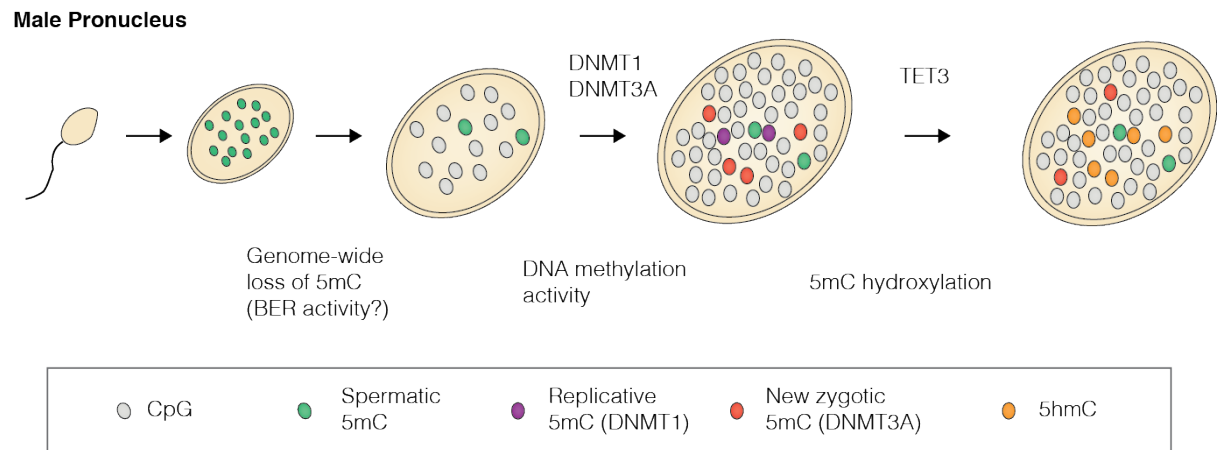


Figure 7. Initial loss of paternal 5mC is independent of TET3. In this model, the initial loss of DNA methylation might be due to base-excision repair (BER) and TET3 protects newly hypomethylated sequences from accumulating new DNA methylation. Modified from (Amouroux et al. 2016).

Recently however, it was shown that ablation of maternal TET3 did not affect early loss of paternal 5mC observed in zygotes. A new model was proposed, in which the initial loss of paternal 5mC does not require 5hmC formation as genetic ablation of TET3, impedes 5hmC accumulation in zygotes without affecting the early loss of paternal 5mC (Figure 7). Instead, the initial loss of 5mC might be due the activity of base-excision repair (BER) pathway (Hajkova et al. 2010; Santos et al. 2013). Further, 5hmC accumulation is dependent on the activity of zygotic DNMT3A and DNMT1, documenting a role for TET3-driven hydroxylation in protecting the newly acquired hypomethylated state from *de novo* DNA methylation (Amouroux et al. 2016).

After the fusion of paternal and the maternal pronuclei, genome-wide 5hmC, 5fC and 5caC in sperm-derived and 5mC in oocyte-derived chromosomes are lost by passive dilution (Inoue and Zhang 2011). However, sequences such as maternally derived methylated CpG-rich regions, including maternal imprinting control regions (ICRs) stay partially or even fully methylated mediated by DNMT1 and DNMT3A (Smallwood et al. 2011; Smith et al. 2012; Xiajun Li et al. 2008).

The regulation of TET expression is dependent on developmental stage and cell type indicating distinct functions of the different members (Szwagierczak et al. 2010; H. Wu and Zhang 2014). The ICM of the blastocyst harbors pluripotent ESCs that express *TET1* and *TET2* at relatively high levels. Although both proteins are dispensable for ESC maintenance, they are important for differentiation into defined lineages (Koh et al. 2011).

TET1 or TET2 mutant mice are viable, while *TET1/TET2* double-knockout mice die perinatally with severe developmental defects, suggesting a role of

Introduction

TET1 and TET2 in regulating embryonic development. TET3 might be partially compensating functions of TET1 and TET2, as some of the double-mutant embryos survived to normal and fertile adult mice (Dawlaty et al. 2011, 2013). Consistent with the important function of TET3 in early embryonic development, *TET3* knockout mice die perinatally (Gu et al. 2011). In brain and other somatic tissues, TET enzymes are expressed, indicating that DNA methylation might be more dynamic as previously thought (H. Wu and Zhang 2014). Interestingly, there is growing evidence for 5hmC to be a stable epigenetic mark rather than only an intermediate in the process of active demethylation. Recently, specific 5hmC readers in different tissues were identified that might be involved in different regulatory processes (Spruijt et al. 2013). Furthermore, TET1 is involved in keeping promoters and enhancers of highly expressed genes as well as CpG islands hypomethylated to enable rapid gene activation upon differentiation (Ficz et al. 2011; Neri et al. 2015).

In summary, the epigenetic reprogramming in the early embryo is a complex process regulated by a dynamic interplay between active DNA demethylation, *de novo* DNA methylation and TET mediated 5mC hydroxylation. Although, many processes are understood, further studies will be necessary to investigate locus-specific modification changes and key factors involved in this process.

1.4 DNA modifications in neurogenesis and neurodegeneration

During embryonic and adult neurogenesis, neuronal stem cells follow a precise path of differentiation and give rise to functional neurons at various developmental stages. This process is highly regulated by epigenetic factors that encompass histone modifications, non-coding regulatory RNAs and DNA modifications. In the next chapter the function of dynamic DNA modifications in embryonic and adult neurogenesis as well as neurodegeneration will be introduced.

1.4.1 DNA modification dynamics during embryonic and adult neurogenesis

During early embryonic development and throughout life, neurogenesis gives rise to functional neurons from neuronal stem cells (NSCs) and progenitor cells (NPCs), which are located in the subgranular zone (SGZ) of the dentate gyrus in the hippocampus and the subventricular zone (SVZ) of the lateral ventricle in the adult brain (Ma et al. 2010). Epigenetic mechanisms interplay in a spatially and temporally regulated manner to regulate this highly complex process (Figure 8) (Santos-Reboucas and Pimentel 2007).

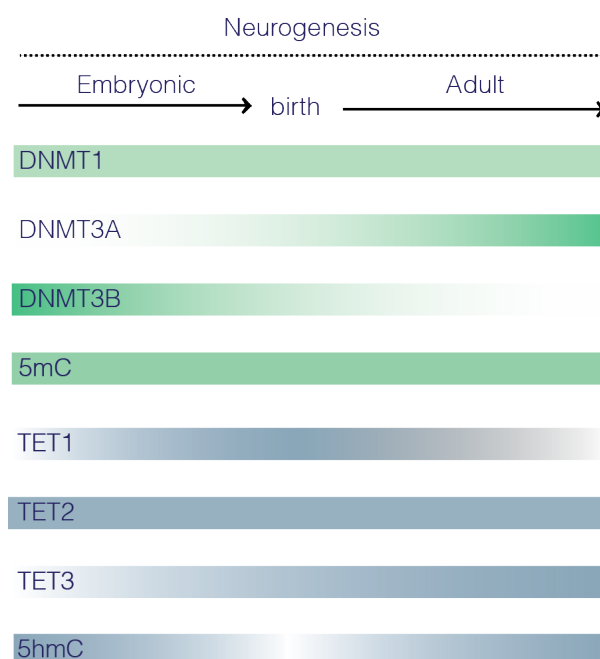


Figure 8. Embryonic and adult neurogenesis. Expression of epigenetic factors and levels of DNA modifications during embryonic and adult neurogenesis. Data from (Ehrlich et al. 1982; Goto et al. 1994; D. Watanabe, Uchiyama, and Hanaoka 2006; Ito et al. 2010; Szlagierczak et al. 2010; Wen et al. 2014; Zhiqin Wang et al. 2016).

DNMT1 is highly expressed in the embryonic and adult brain especially in postmitotic neurons (Goto et al. 1994). Consistently, DNA methylation by DNMT1 is essential for brain development and neuronal maturation (Fan et al. 2001). Interestingly, neurons have high levels of DNA methylation (Ehrlich et al. 1982)

Introduction

and the majority of the DNMT1 is localized in the cytoplasm and enzymatically active (Inano et al. 2000) hinting at other functions of DNMT1 besides DNA methylation. Depletion of DNMT1 in neurons is not fatal, but it was suggested that restoration of DNA methylation after BER of G-T mismatches is mediated by DNMT1 (Brooks, Marietta, and Goldman 1996).

The *de novo* methyltransferases show an opposite expression profile with a transition from DNMT3B to DNMT3A expression during neuronal development. DNMT3B regulates early embryogenesis and is highly expressed in NPCs, while DNMT3A is highly expressed postnatally decreasing towards adulthood and was further shown to establish tissue-specific methylation patterns in neurons (Daisuke Watanabe et al. 2002, 2004; Feng et al. 2005; D. Watanabe, Uchiyama, and Hanaoka 2006).

DNMT3A influences the neuronal and glial cell differentiation of embryonic NSC *in vitro* as ESCs depleted of DNMT3A differentiate into astrocytes and oligodendrocytes and show other defects such as hypomethylation and a higher proliferation rate (Z. Wu et al. 2012). Furthermore, DNA methylation patterns set by DNMT3A on intergenic regions and gene bodies antagonize PcG mediated repression of neurogenic genes and thereby maintaining an active chromatin state of genes that are important for development (H. Wu et al. 2010).

The paralogue DNMT3B has a role in regulating the timing of embryonic neural differentiation and maturation as RNAi mediated DNMT3B knockdown experiments in human ESCs showed premature maturation, neural and neural crest marker expression (Martins-Taylor et al. 2012).

Interestingly, in brain tissues like cortex and hippocampus, 5hmC abundance is among the highest in comparison to other tissues (Szwagierczak et al. 2010) pointing towards a possible functional importance of this DNA modification in neuronal development and activity. The 5hmC levels increase during neuronal differentiation and all TET enzymes are expressed in the brain with TET2 and TET3 showing the highest levels (Figure 8) (Ito et al. 2010; Szwagierczak et al. 2010; C.-X. Song et al. 2011; Szulwach et al. 2011; Wen et al. 2014).

Interestingly, 5hmC is not enriched at enhancers but associates preferentially with gene bodies of activated neuronal function-related genes. Within these genes, 5hmC is accompanied by loss of H3K27me3. Loss of function of TET2/3 and gain of function of EZH2 leads to defects in neuronal differentiation, suggesting that formation of 5hmC and loss of H3K27me3 cooperate to promote brain development (Hahn et al. 2013).

TET3 has been associated with fear extinction learning by triggering a redistribution and accumulation of 5hmC, which is required for rapid behavioral adaptation (Xiang Li et al. 2014). Although, TET1 knockout mice are viable (see

1.3.2), spatial learning and memory is affected in these animals. Further, several genes of NSCs are hypermethylated and thus silenced indicating an important role of TET1 in differentiating NPCs to neurons (R.-R. Zhang et al. 2013).

In summary, the DNMT and TET family carry out diverse functions in regulating specific aspects of neurogenesis. The deregulation of epigenetic modifications as DNA methylation is associated with neurodegenerative diseases and will be introduced in the next part.

1.4.2 DNA methylation and neurodegenerative diseases

Human diseases have been associated with alterations in epigenetic mechanisms including cancer, syndromes associated with chromosomal instability, imprinting and neurodegenerative disorders as well as mental retardation. In contrast to the genetic information, epigenetic marks are reversible, which makes them a possible target for disease treatment (Santos-Reboucas and Pimentel 2007; Pogribny and Beland 2009).

Given the critical functions of DNMTs mentioned in the previous section, it is not surprising that misregulation of DNA methylation was described to be involved in the pathophysiology of neurodegenerative disorders (Zhao, Deng, and Gage 2008; Braun and Jessberger 2014). In several studies, neurological disorders were described to be caused by mutations within DNMTs (Table 1).

Table 1. DNMT family members and related neurological diseases or defects

Gene	Molecular defect	Related disease/ phenotype	References
<i>DNMT1</i>	N-terminal TS domain mutations; heterozygous C-terminal TS domain mutations; heterozygous	Hereditary sensory and autonomic neuropathy type 1 (HSAN1E); Autosomal dominant cerebellar ataxia, deafness, and narcolepsy (ADCA-DN)	(Klein et al. 2011; Winkelmann et al. 2012)
<i>DNMT3A</i>	DNMT3A functional mutations/ deletions	Acquisition of developmental mental defects; impaired postnatal neurogenesis	(Nguyen et al. 2007; H. Wu et al. 2010; Yao and Jin 2014)
<i>DNMT3B</i>	DNMT3B functional mutations; usually heterozygous in catalytic domain	The immunodeficiency, centromere instability, facial anomalies (ICF) syndrome	(Hansen et al. 1999; G. L. Xu et al. 1999; Robertson 2005)

Introduction

To date, DNMT3A has not been associated with neuronal diseases, although mice depleted of functional DNMT3A in the central nervous system (CNS) die prematurely due to developmental defects such as hypoactivity and defective motor coordination (Nguyen et al. 2007). Further, DNMT3A deletion leads to a postnatal neurogenesis defect in both the SVZ and SGZ as significantly fewer terminally differentiated neurons arise from *DNMT3A*^{-/-} NSCs indicating that DNMT3A regulates postnatal neurogenesis (H. Wu et al. 2010).

Polymorphic mutations in the *de novo* methyltransferase DNMT3B are associated with a syndrome called immunodeficiency, centromere instability, and facial anomalies (ICF) (Hansen et al. 1999; G. L. Xu et al. 1999). These patients suffer from immunodeficiency, which results from a significant reduction of at least two immunoglobulin isotypes. Further, ICF patients show hypomethylation and decondensation of pericentromeric heterochromatin on chromosomes 1, 16 and 9 in mitogen stimulated B cells or lymphoblastoid cell lines (Robertson 2005).

Heterozygous DNMT1 mutations causing DNA hypomethylation in patients suffering from hereditary sensory and autonomic neuropathy type 1E (HSAN1E) or autosomal dominant cerebellar ataxia deafness and narcolepsy (ADCA-DN) have been reported by several medical studies. Clinical manifestations of HSAN1E and ADCA-DN include a broad phenotypic spectrum with sensory impairment, hearing loss and dementia as hallmarks of the disease (Baets et al. 2015). In particular, HSAN1E is characterized by late onset neurologic disorders of the CNS and the peripheral nervous system (PNS) that manifest in progressive degeneration predominantly of sensory and autonomic neurons (Klein et al. 2013). Notably, all mutations described to date affect the TS domain in the regulatory part of DNMT1 (Klein et al. 2011; Winkelmann et al. 2012; Pedroso et al. 2013; Yuan et al. 2013; Moghadam et al. 2014; Z. Sun et al. 2014; Baets et al. 2015), raising the question how simple amino acid exchanges outside the catalytic domain of DNMT1 may lead to global hypomethylation and late onset neurodegenerative diseases. The pathophysiological trajectories underlying these neurological disorder caused by DNMT1 mutations remain mostly unknown.

1.5 New advances in genome engineering techniques

To gain insights into the function of a gene a widely used approach remains targeted disruption and the subsequent assay for cellular defects in living cells. Genome editing enables functional studies of genetic variation in biology and disease, and holds a great potential for clinical applications. Zinc-Finger Nucleases (ZFNs) and Transcription Activator-Like Effector Nucleases (TALENs) have been developed to facilitate the modification of endogenous genomic sequences (Miller et al. 2007; Hockemeyer et al. 2011). However, the applicability of these technologies is limited and laborious, as for each target site a new complex nuclease needs to be designed.

Recently, a powerful tool to edit genomes of diverse organisms revolutionized the genome engineering field by providing an easy and fast method. This approach utilizes the CRISPR (clustered regularly interspaced short palindromic repeats)/Cas system, which was initially discovery as an adaptive immune system in bacteria. In general, the CRISPR-associated protein 9 (Cas9) is an RNA-guided DNA endonuclease that can be recruited to sites in the genome by editing the guide RNA (gRNA) sequence (Figure 9) (Haurwitz et al. 2010; Mali, Esvelt, and Church 2013). The gRNA mediates the specificity of Cas9 in the genome by binding 20 bp within the target sequence and thereby recruits endonuclease Cas9, which in turn introduces a DNA double strand break (DSB).

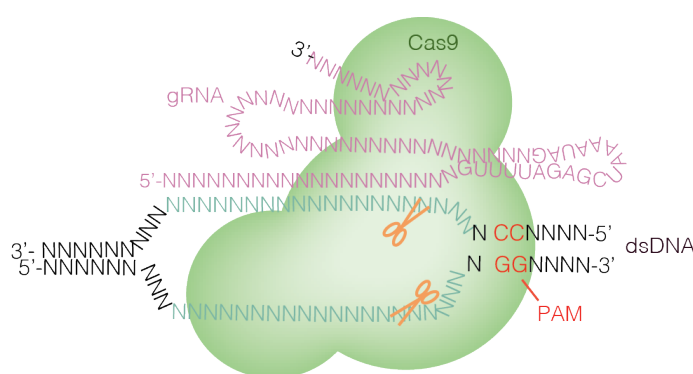


Figure 9. The Cas9-gRNA RNA-guided nuclease complex for eukaryotic genome engineering. Target recognition and cleavage require a 20 bp sequence complementary to the genomic target sequence and presence of the appropriate NGG PAM (orange) sequence. Modified from (Mali, Esvelt, and Church 2013).

DSBs are either repaired by highly efficient yet error-prone non-homologous end joining (NHEJ) or by inefficient homology directed repair (HDR). While NHEJ is a useful tool for gene knockout generation, it is not desirable for precise modification of the genome. In contrast, HDR allows targeted manipulation of the

Introduction

genome, but competes with highly active NHEJ (Capecchi 2005). Although this approach has improved the generation of gene disruptions in diverse organisms, off-target effects have to be considered (Kuscu et al. 2014).

In addition to the CRISPR/Cas9 system, phage-derived serine integrases, in particular BXB1, were used as novel genome engineering tools as they are unidirectional and highly efficient site-specific recombinases (Z. Xu et al. 2013). In general, the serine integrases induce the recombination of a phage attachment sites (attP) and bacterial attachment sites (attB) (Brown et al. 2011), which can facilitate in-frame insertions of desired sequences such as prefabricated functional cassettes.

In this study, we aim to combine the advantages of both CRISPR/Cas and unidirectional integrases into one fast, widely applicable and flexible method.

1.6 New advances in transcriptome analysis

As described before (1.5), targeted gene disruption by CRISPR/Cas-based approaches and the subsequent assay for cellular defects in living cells is used to investigate gene function. The function of the desired gene disruption on transcriptional regulation, can be investigated by RNA sequencing (RNA-seq), a genome-wide quantification of mRNA transcripts.

The appreciation for cell-to-cell variability and the ability to profile the transcriptome of individual cells, has transformed the characterization of cell types, differentiation states and rare cellular phenotypes, revealing exciting biological and medical insights (Sandberg 2014; Kolodziejczyk et al. 2015; Ebinger et al. 2016).

A wide variety of single cell RNA-sequencing (scRNA-seq) methods, such as Smart-seq, automated microfluidic platform from Fluidigm (C1 platform) and CEL-seq, have been established (Hashimshony et al. 2012; Ramsköld et al. 2012; Picelli et al. 2013). Therefore, it is important to compare different parameters, in order to choose the best method for a given question. Some strengths and weaknesses concerning monetary costs, sensitivity, accuracy and coverage of different methods are already known (Hashimshony et al. 2012; A. R. Wu et al. 2014). However, a thorough and systematic comparison of relevant parameters across scRNA-seq methods is still lacking.

1.7 Aims of this work

The main questions addressed in this PhD thesis revolve around the functional and regulatory interplay between epigenetic key factors DNMT1 and UHRF1, its role during development and disease as well as technology advancements in genome editing and transcriptomics towards a better understanding of chromatin biology.

DNMT1 TS domain mutations are linked to a neurodegenerative disease called HSAN1E, however, the underlying molecular dysfunctions remained elusive. Here, I aimed at characterizing two disease related mutations in mouse DNMT1 (P496Y and Y500C) with respect to their impact on catalytic activity, UHRF1 interaction, subnuclear localization, protein stability and neuronal progenitor differentiation.

Although, it is well established that UHRF1 is essential for maintenance DNA methylation, defined functions of distinct regulatory domains remained unknown. Therefore, the next objective of this work was to investigate UHRF1-mediated modifications and their functional interplay with DNMT1, regulating DNA methylation. Besides chromatin binding domains, UHRF1 carries a RING domain that ubiquitinates histones. With proteomic analysis, we aimed at identifying novel, site-specific histone ubiquitination set by UHRF1. As the identified UHRF1-dependent H3 ubiquitination was essential for DNMT1 targeting and function, I participated in further analyzing the underlying regulatory mechanisms and regions in DNMT1 responsible for ubiquitinated histone binding.

To further study the E3 ligase function of UHRF1 and UHRF2, I co-established a multifunctional integrase (MIN) tag for rapid and versatile genome engineering. The MIN tag serves as a genetic entry site for functional as well as knockout modules. Based on CRISPR/Cas mediated manipulation, this approach enables efficient generation of multiple isogenic cell lines to study gene function under physiological conditions or gene depletion.

Further, I was interested in non-histone ubiquitination targets of UHRF1 and UHRF2. Using the MIN tag strategy, I established ESCs depleted of either UHRF1 or UHRF2 and assayed for differential ubiquitination in living cells with mass spectrometry. To exclude redundancy between the two paralogues UHRF1 and UHRF2, which are highly similar in both sequence and structure, I compared the ubiquitome of *UHRF1* and *UHRF2* knockout cells and screened for novel ubiquitination targets.

New next-generation technologies to profile the transcriptome of individual cells now enable researchers to investigate exciting biological and medical

questions. In the cancer evolution field, cell-to-cell variability is of special interest as it is a major driver of cancer evolution, progression, and emergence of drug resistance. To investigate cell heterogeneity, a wide variety of scRNA-seq methods have been established. In this project, I participated in comparing different library preparation methods, in order to systematically compare parameters across scRNA-seq methods.

Results

2 Results

2.1 DNMT1 mutations found in HSAN1E patients affect interaction with UHRF1 and neuronal differentiation

ORIGINAL ARTICLE

DNMT1 mutations found in HSNIE patients affect interaction with UHRF1 and neuronal differentiation

Martha Smets[†], Stephanie Link^{†,‡}, Patricia Wolf[¶], Katrin Schneider[§],
Veronica Solis^{||}, Joel Ryan, Daniela Meilinger, Weihua Qin and
Heinrich Leonhardt*

Department of Biology II and Center for Integrated Protein Science Munich (CIPSM), Ludwig-Maximilians-Universität München, 82152 Planegg-Martinsried, Germany

*To whom correspondence should be addressed. Tel: +49 89 218074232; Fax: +49 89 218074236; Email: h.leonhardt@lmu.de

Abstract

DNMT1 is recruited to substrate sites by PCNA and UHRF1 to maintain DNA methylation after replication. The cell cycle dependent recruitment of DNMT1 is mediated by the PCNA-binding domain (PBD) and the targeting sequence (TS) within the N-terminal regulatory domain. The TS domain was found to be mutated in patients suffering from hereditary sensory and autonomic neuropathies with dementia and hearing loss (HSANIE) and autosomal dominant cerebellar ataxia deafness and narcolepsy (ADCA-DN) and is associated with global hypomethylation and site specific hypermethylation. With functional complementation assays in mouse embryonic stem cells, we showed that DNMT1 mutations P496Y and Y500C identified in HSNIE patients not only impair DNMT1 heterochromatin association, but also UHRF1 interaction resulting in hypomethylation. Similar DNA methylation defects were observed when DNMT1 interacting domains in UHRF1, the UBL and the SRA domain, were deleted. With cell-based assays, we could show that HSNIE associated mutations perturb DNMT1 heterochromatin association and catalytic complex formation at methylation sites and decrease protein stability in late S and G2 phase. To investigate the neuronal phenotype of HSNIE mutations, we performed DNMT1 rescue assays and could show that cells expressing mutated DNMT1 were prone to apoptosis and failed to differentiate into neuronal lineage. Our results provide insights into the molecular basis of DNMT1 dysfunction in HSNIE patients and emphasize the importance of the TS domain in the regulation of DNA methylation in pluripotent and differentiating cells.

[†]Present address: BioMedical Center (BMC), Department of Molecular Biology, Ludwig-Maximilians-Universität München, Großhaderner Str. 9, 82152 Planegg-Martinsried, Germany.

[‡]Present address: Microcoat Biotechnologie GmbH, Am Neuland 3, 82347 Bernried am Starnberger See, Germany.

[§]Present address: Definiens AG, Bernhard-Wicki-Str. 5, 80636 München, Germany.

[¶]Present address: Max Planck Institute of Neurobiology, Am Klopferspitz 18, 82152 Martinsried, Germany.

^{||}The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

Received: December 21, 2016. Revised: February 7, 2017. Accepted: February 9, 2017

© The Author 2017. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Introduction

Epigenetic mechanisms are crucial for the regulation of gene expression during embryonic development and cell differentiation. Tissue-specific DNA methylation patterns are established during embryogenesis by the *de novo* DNA methyltransferases DNMT3A and DNMT3B, whereas the propagation of these marks to future somatic cell generations is based on the maintenance DNA methyltransferase 1 (DNMT1) (1–5). The catalytic activity of DNMT1 is attributed to its C-terminal domain (CTD), however enzyme regulation, targeting and activation are mediated by the N-terminal domain (NTD) harboring distinct subdomains (6). During S phase DNMT1 localization at sites of DNA replication is mediated by the PCNA-binding domain (PBD) while heterochromatin binding during late S and G2 is mediated by the targeting sequence (TS) domain, both of which contribute to proper maintenance of DNA methylation patterns (7–9).

A key factor in the regulation of DNMT1 is Ubiquitin-like, containing PHD and RING finger domains 1 (UHRF1, also known as 95 kDa mouse nuclear protein (Np95)). By binding to hemimethylated DNA via its SET and RING associated (SRA) domain (10–14), UHRF1 targets DNMT1 to its substrate sites (10). In addition, UHRF1 binds to methylated H3K9 via its tandem Tudor domain (TTD) and to H3R2 via its plant homeodomain (PHD) (15–18).

By cooperative binding of repressive H3K9me3 marks and hemimethylated DNA, UHRF1 targets DNMT1 to newly synthesized DNA in heterochromatin after replication (19). In addition, the UHRF1 RING domain ubiquitinates H3 tails on K18 (K23 in *Xenopus*), which is specifically recognized by the ubiquitin interacting motif (UIM) in the TS domain of DNMT1 and required for DNA methylation (20,21). Besides intermolecular protein–protein binding, also intramolecular protein interactions serve as a prerequisite for DNMT1 activation. Firstly, in complex with unmethylated DNA, the linker between the zinc finger (CXXC) domain and the bromo-adjacent homology domain 1 (BAH1) blocks the access of DNA to the catalytic center (22). Secondly, the crystal structure of DNMT1 reveals that, in absence of DNA, the TS domain is inserted in the DNA-binding pocket of the CTD thereby inhibiting enzymatic activity (23). These two autoinhibitory mechanisms have to be overcome by structural changes before the methylation reaction can occur. Interaction of UHRF1 with DNMT1 releases the TS domain and enables catalytic activity of the CTD (24).

In addition to enzyme targeting and activation, protein stability also contributes to the regulation of maintenance DNA methylation. Stability and abundance of DNMT1 during the cell cycle is governed by UHRF1 dependent ubiquitination and deubiquitination by the ubiquitin specific peptidase 7 (USP7, also known as herpes virus associated ubiquitin specific protease (HAUSP)) which protects against proteasomal degradation (25,26). While Tip60 mediated acetylation promotes ubiquitination by UHRF1 and thereby marks DNMT1 for proteasomal degradation, the corresponding deacetylation by histone deacetylase 1 (HDAC1) contributes to the stabilization of DNMT1 (25,26).

Despite its well-known replication-coupled function as maintenance DNA methyltransferase in proliferating cells, DNMT1 is highly expressed in embryonic and adult postmitotic neurons, especially in the central nervous system (CNS) (27,28). Remarkably, DNA methylation is required in adult neurogenesis and its misregulation was described to be involved in the pathophysiology of neurodegenerative disorders (29,30). Several medical studies have reported heterozygous DNMT1 mutations causing DNA hypomethylation in patients suffering from hereditary sensory and autonomic neuropathy type 1E (HSANIE, OMIM 614116) or autosomal

dominant cerebellar ataxia deafness and narcolepsy (ADCA-DN, OMIM 604121). Strikingly, all causative mutations described to date affect a genomic region in DNMT1, which codes for the TS domain (31–38). Clinical manifestations of HSANIE and ADCA-DN include a broad phenotypic spectrum with sensory impairment, hearing loss and dementia as hallmarks of the disease (38). In particular, HSANIE is characterized by late onset neurologic disorders of the CNS and the peripheral nervous system that manifest in progressive degeneration predominantly of sensory and autonomic neurons (39). However, the pathophysiological trajectories underlying these neurological disorders caused by DNMT1 mutations remain mostly unknown.

In this study, we investigate the effect of DNMT1 mutations identified in HSANIE patients on the function of the TS domain in embryonic stem cells and neuronal progenitor cell differentiation. With functional complementation assays, we show that HSANIE associated mutations in mouse DNMT1 disrupt binding of UHRF1 and lead to defects in maintenance DNA methylation. We show that the TS point mutations affect the association with late replicating chromatin, catalytic complex formation at methylation sites and cell cycle dependent protein stability.

Importantly, ESCs expressing HSANIE mutants fail to differentiate into the neuronal lineage and are prone to undergo apoptosis. These results emphasize the importance of the complex interplay of UHRF1 with the TS domain in regulating DNMT1 function and DNA methylation in neurogenesis.

Results

Deletions and TS domain mutations found in HSANIE patients affect DNMT1 activity *in vivo*

Although the catalytic activity of DNMT1 resides in its CTD, the NTD of the enzyme is indispensable for proper maintenance DNA methylation (6). Within the NTD the TS domain spans from amino acid 309 to 628 in mouse isoform 2 and is highly conserved among different species (Fig. 1). To map regions required for methylation activity, we generated a systematic set of DNMT1 deletion constructs and performed rescue experiments expressing the deletion constructs in *Dnmt1*^{−/−} mouse embryonic stem cells (ESCs). Site specific DNA methylation analysis revealed that not only the regions between amino acid (aa) 356 to 404 and 458 to 500 (21), but also regions from 496 to 573 are functionally relevant to maintain methylation patterns at pericentromeric heterochromatin (Fig. 2A, shown in dark blue). DNMT1 either lacking the flexible, less conserved N-terminal region of the TS domain (aa 309 to 355), or the C-terminal region of the TS domain (aa 579 to 595) was able to restore methylation at major satellite repeats (Fig. 2A, shown in light blue).

Interestingly, mutations of two highly conserved amino acids in the TS domain (Fig. 1B) were linked to a neurodegenerative disease described as HSANIE (31,39). As the HSANIE associated mutations are located within a functionally relevant central region of the TS domain (Fig. 2A), we investigated the DNA methylation activity of GFP-DNMT1 harboring HSANIE mutations. To this end, we cloned HSANIE associated mutations P496Y (human: D490E.P491Y) and Y500C (human: Y495C) in mouse DNMT1 expression constructs and reintroduced the mutant proteins in *Dnmt1*^{−/−} ESCs (Supplementary Material, Fig. S6). In line with previous findings (31), we found that transiently and stably expressed mutated DNMT1 is unable to rescue DNA methylation levels *in vivo* (Fig. 2B, Supplementary Material, Fig. S1A and S1C). Methylation levels at imprinted and unmethylated H19 promoter remained unchanged (Supplementary Material, Fig. S1B).

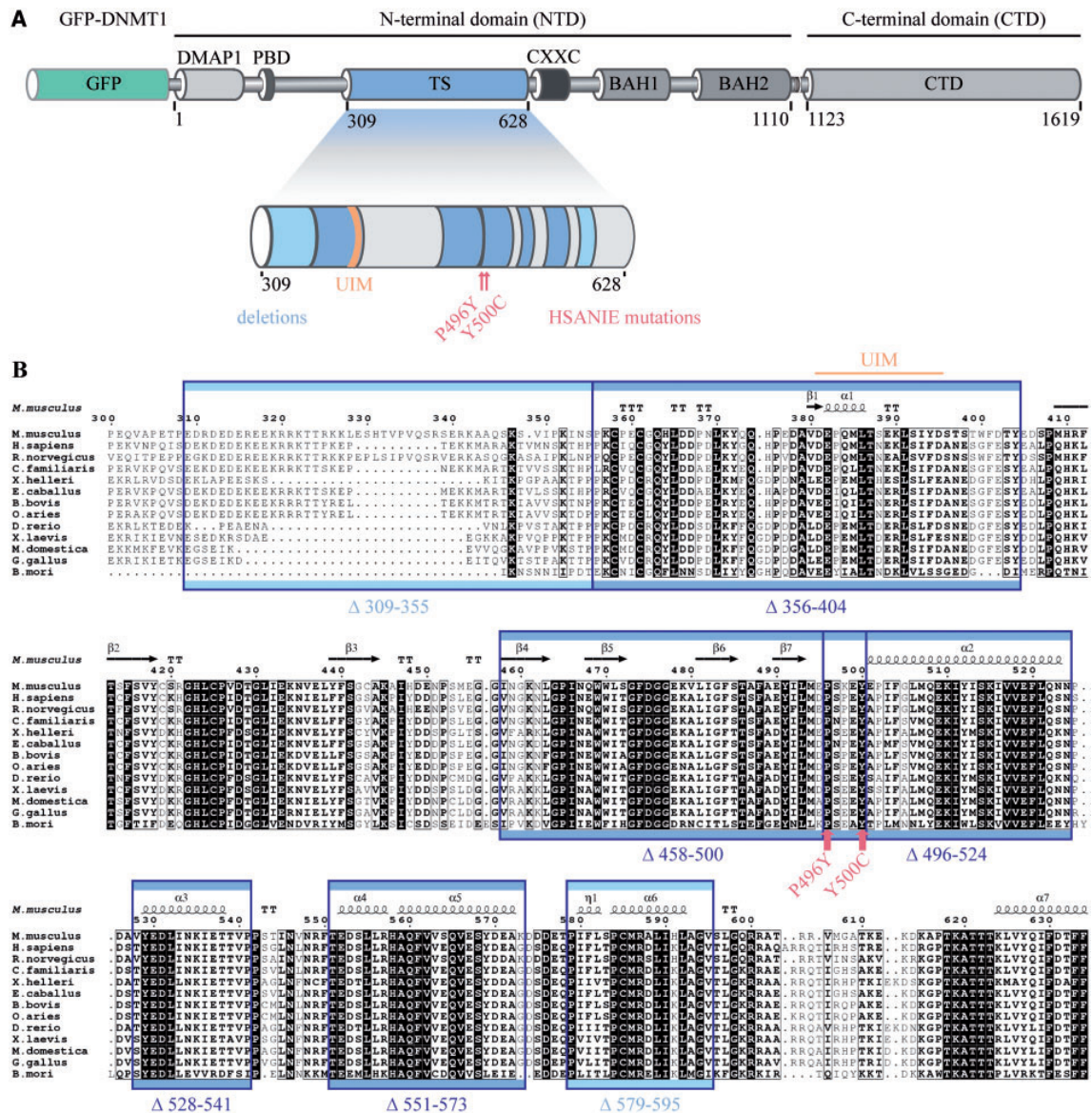


Figure 1. Domain structure of mouse DNMT1 and conservation of the TS domain among different species. (A) Domain structure of GFP-DNA methyltransferase 1 (DNMT1) with illustration of deletions and HSNIE associated point mutations in the targeting sequence (TS) domain. The large regulatory N-terminal domain (NTD) of DNMT1 is comprised of a DNA methyltransferase associated protein 1 (DMP1)-binding domain, a proliferating cell nuclear antigen (PCNA)-binding domain (PBD), the TS domain harboring a ubiquitin interacting motif (UIM (21)), a zinc finger (CXXC) domain, two bromo-adjacent homology (BAH1 and BAH2) domains and a catalytically active C-terminal domain (CTD). (B) Primary sequence alignment of TS domains from different species. The secondary structure of the mouse TS domain is indicated (PDB: 3AV4 (23)). Highly conserved residues are shaded black. Deleted regions are indicated by blue rectangles and hereditary sensory and autonomic neuropathy type IE (HSNIE) associated point mutations by red arrows.

To investigate a potential targeting defect of DNMT1 harboring HSNIE associated mutations, we made use of a DNA methyltransferase trapping assay in living cells. In this assay, the cytosine analogue 5-aza-2'-deoxycytidine (5-aza-dC) is incorporated into DNA during replication and forms an irreversible covalent complex with the active methyltransferase at the C6 position of the cytosine residue thereby immobilizing the catalytically active enzyme at DNA replication foci (40). Trapping was measured by fluorescence recovery after photobleaching (FRAP). We found that DNMT1 harboring HSNIE mutations can be trapped, but exhibit a larger mobile fraction than DNMT1 wild-type (wt) indicating a lower targeting to substrate sites in living cells (Fig. 2C). In

summary, we could show that the TS domain of DNMT1 is essential for maintenance DNA methylation. Deletions in the TS domain between amino acid 356 to 573 as well as point mutations found in HSNIE patients lead to decreased methylation activity *in vivo*.

Functional relevance of UHRF1 domains on the regulation of DNMT1 enzymatic activity

UHRF1 is required for maintaining DNA methylation patterns after replication by direct interaction with DNMT1 (10). To elucidate which protein domains are involved in the interaction of

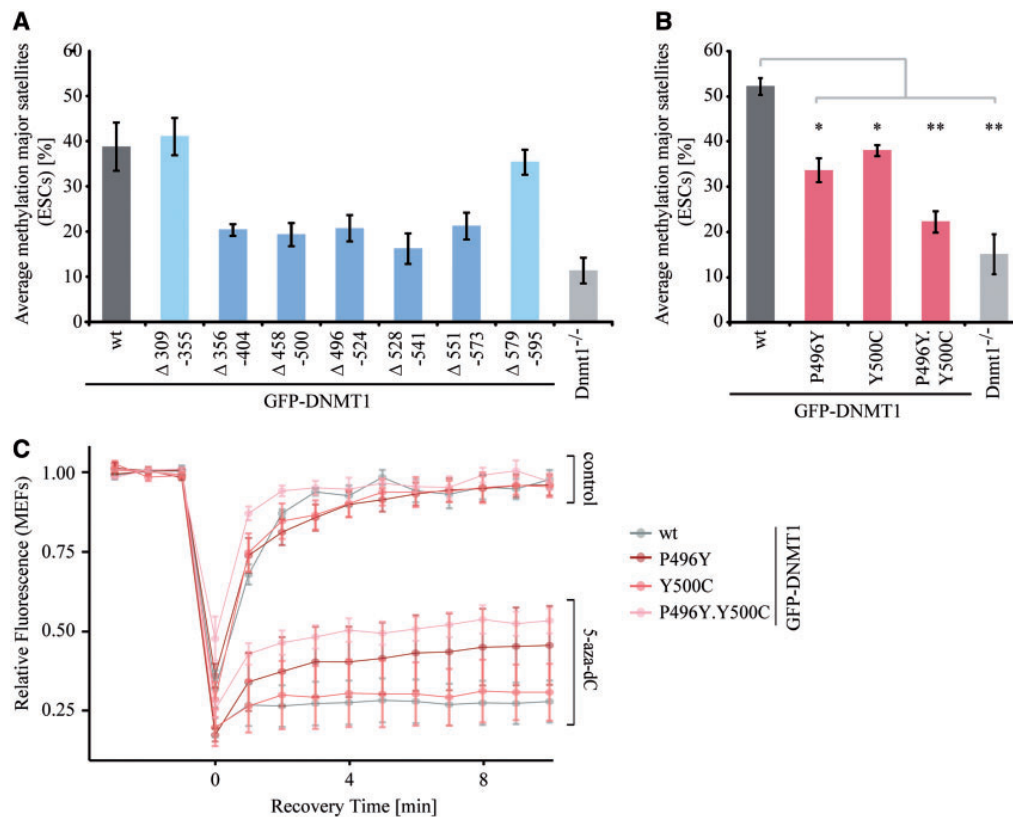


Figure 2. TS domain deletions and HSNIE associated mutations affect the DNA methylation activity of DNMT1 in vivo. (A and B) Site-specific DNA methylation analysis at major satellite repeats of mouse *Dnmt1*^{-/-} ESCs expressing transiently GFP-DNMT1 wild-type (wt) and deletions (A) or HSNIE associated point mutant constructs (B). DNA methylation levels of untransfected *Dnmt1*^{-/-} ESCs are displayed for comparison. Shown are mean values \pm standard deviations (SDs) from two to six (A) or (B) three independent biological replicates (average from eight CpG sites, respectively). (B) Two-sample t-tests were performed that assume equal variances. Asterisks represent significant differences: * $P < 0.05$, ** $P < 0.001$. (C) Quantitative evaluation of FRAP data showing mean curves. Error bars represent standard deviations. FRAP of S-phase mouse embryonic fibroblasts ($n = 5$) expressing GFP-DNMT1 and GFP-DNMT1 harboring HSNIE mutations without drug treatment (upper panel) and after treatment with 30 μ M 5-aza-dC for 30 min (lower panel).

DNMT1 and UHRF1 (Fig. 3A), we generated a systematic set of constructs of the NTD and CTD of DNMT1 as well as single domains of UHRF1 and performed a binding assay. We used a semi-quantitative fluorescence protein-protein interaction assay, in which GFP-tagged proteins are immobilized on a GFP-multiTrap plate, incubated with RFP/Ch-tagged proteins and binding ratios are determined by fluorescence readout (41). Mapping studies on UHRF1 showed that the SRA domain interacted with DNMT1 (Fig. 3C). In terms of DNMT1, the TS mediated the binding to UHRF1 (Fig. 3B). To analyze domain contribution to the TS domain interaction in the UHRF1 protein context, we tested the binding of Ch-TS to UHRF1-GFP single domain deletions (Fig. 3D). We generated constructs with single domain deletions of UHRF1 and found that deletions of the TTD, the PHD and the RING domain had little to no effect on the interaction, whereas deletions of the UBL and the SRA domain reduced the binding of Ch-TS to UHRF1-GFP indicating that the TS domain of DNMT1 can interact with two domains of UHRF1, the UBL and the SRA domain of UHRF1 (Fig. 3F). In our experimental set-up, the UBL domain alone, however, did not interact with DNMT1 suggesting a cooperative binding mode involving the UBL and SRA domain of UHRF1.

To examine which domains of UHRF1 are functionally relevant for regulation of DNMT1 enzymatic activity, we

investigated the ability of UHRF1 single domain deletions in mediating DNA methylation in vivo. All *Uhrf1*^{-/-} ESC lines stably expressing UHRF1-GFP with single domain deletions displayed low methylation levels suggesting a DNMT1 recruitment defect (Fig. 3E). Consequently, we investigated the dependence of DNMT1 subnuclear localization on UHRF1 domains. In the wt UHRF1-GFP cell line, DNMT1 showed late S phase-specific heterochromatin association that was abolished in all single domain UHRF1 deletion cell lines displaying diffuse nuclear localization of DNMT1 comparable to *Uhrf1*^{-/-} ESCs (Supplementary Material, Figs S3 and S4). In conclusion, all UHRF1 domains are required for recruitment of DNMT1 for maintenance DNA methylation.

HSNIE associated mutations in TS domain of DNMT1 affect the interaction with UHRF1 and lead to dissociation from heterochromatin

As HSNIE associated mutations in the TS domain of DNMT1 lead to hypomethylation at pericentric heterochromatin and the TS mediates the interaction with UHRF1, we investigated the effects of HSNIE mutations on UHRF1 interaction. With coimmunoprecipitation experiments, we found that mutated TS was defective in binding UHRF1 (Fig. 4A), which was

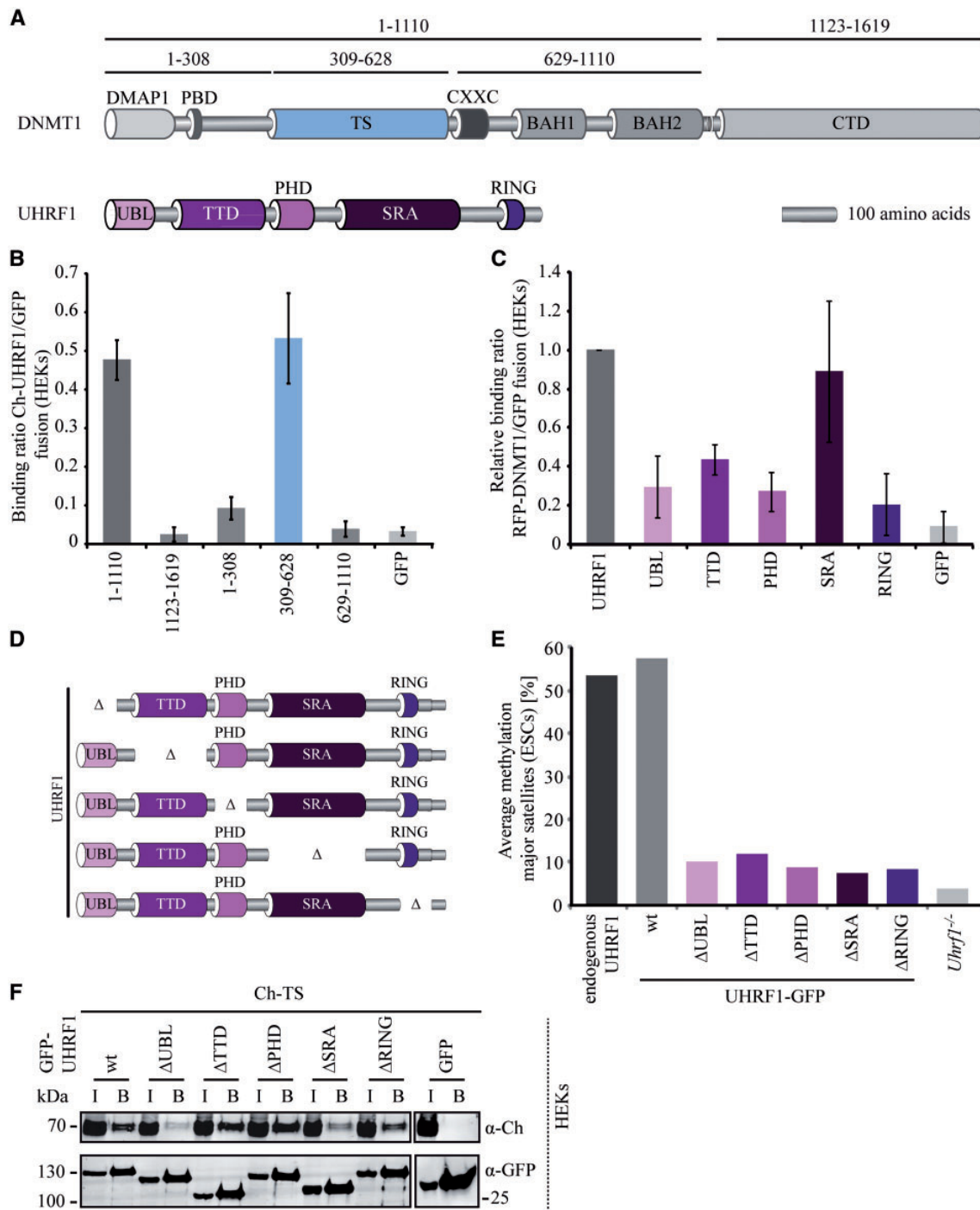


Figure 3. DNMT1 TS domain interaction with UHRF1 is necessary for DNA methylation. (A) Schematic outline of DNMT1 and the ubiquitin-like, containing PHD and RING finger domains 1 (UHRF1) expression constructs used for protein–protein interaction mapping studies. UHRF1 harbors an ubiquitin-like domain (UBL) followed by a tandem Tudor domain (TTD), a plant homeodomain (PHD), a SET and RING associated (SRA) domain and a really interesting new gene (RING) domain. (B) Mapping and relative quantification of the interaction GFP-DNMT1 with Ch-UHRF1 by fluorescence protein–protein interaction assay *in vitro*. Ratios of Ch-UHRF1 over GFP fusion proteins are shown as mean values \pm standard error of the mean (SEM) of three to six biological replicates. (C) Mapping and relative quantification of the interaction GFP-UHRF1 with RFP-DNMT1 by fluorescence protein–protein interaction assay *in vitro*. Ratios of RFP-DNMT1 over GFP fusion proteins are shown as mean values \pm SEM of three biological replicates normalized to the binding ratio of the GFP-UHRF1 full length protein. (D) Schematic outline of different UHRF1-GFP single domain deletion (Δ) constructs used for rescue experiments. (E) Local methylation analysis at major satellite repeats. CpG methylation levels in *Uhrf1*^{-/-} ESCs stably expressing UHRF1-GFP wt or single domain deletions were analyzed. DNA methylation levels of untransfected *Uhrf1*^{-/-} ESCs are shown for comparison. Values represent means from eight CpG sites. (F) Coimmunoprecipitation of UHRF1-GFP wt and single domain deletion mutants and Ch-TS. GFP and Ch fusion constructs were coexpressed in HEK293T cells and cell lysates were used for immunoprecipitation with the GFP-Trap. Bound fractions were detected by immunoblotting with anti-GFP and anti-Ch antibodies. GFP was used as negative control. I = Input, B = Bound.

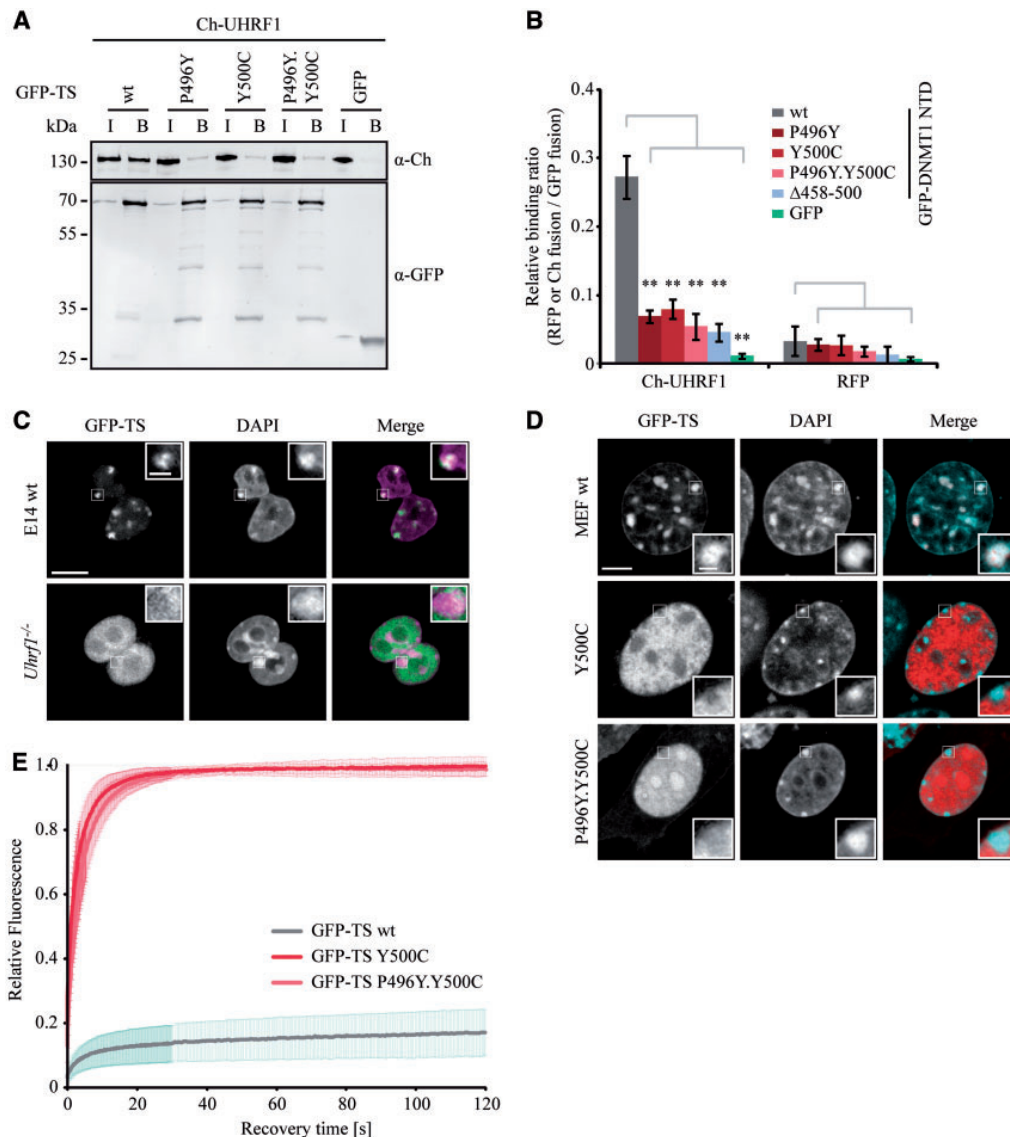


Figure 4. HSNIE associated point mutations in the TS domain reduce the interaction of DNMT1 with UHRF1. (A) Coimmunoprecipitation of GFP-TS wt, P496Y, Y500C and P496Y.Y500C mutant constructs and Ch-UHRF1. GFP and Ch fusion constructs were coexpressed in HEK293T cells and cell lysates were used for immunoprecipitation with the GFP-Trap. Bound fractions were analyzed by immunoblotting with anti-GFP and anti-Ch antibodies. GFP was used as negative control. I=Input, B=Bound. (B) Fluorescence protein-protein interaction assay of the GFP-DNMT1 NTD with Ch-UHRF1. After one-step purification of the GFP-DNMT1 NTD wt and mutant constructs in a GFP-multiTrap plate, the binding of Ch-UHRF1 expressed in HEK293T cells was determined by fluorescence readout. GFP and RFP were used as negative control. Shown are mean relative binding ratios \pm SEM of Ch-UHRF1 or RFP over GFP fusion proteins from four to six biological replicates. Two-sample t-tests were performed that assume equal variances. Significance compared to the relative binding ratio of GFP-DNMT1 NTD wt are indicated: * $P < 0.05$, ** $P < 0.001$. (C) Confocal midsections of fixed mouse ESCs (wt, *Uhrf1*^{-/-}) transiently expressing GFP-TS and DNA was counterstained with DAPI. In the merged image, DAPI is depicted in magenta. Scale bar 10 μ m; enlargements: 3-times magnification, scale bar 2 μ m. (D) Confocal midsections of fixed MEF cells transfected with GFP-TS wt or GFP-TS Y500C and P496Y.Y500C constructs. In the merged image, GFP-TS is depicted in red and DAPI in magenta. Scale bar 5 μ m; enlargements: 3-times magnification, scale bar 1 μ m. (E) Protein mobility of GFP-TS wt and HSNIE associated GFP-TS Y500C and P496Y.Y500C mutants in living MEF cells ($n = 13$) determined by half nucleus fluorescent recovery after photobleaching (FRAP) analysis. Curves represent mean \pm SEM.

confirmed by a semiquantitative fluorescence protein-protein interaction assay (Fig. 4B). Expression of GFP-TS in ESCs revealed that while the TS domain was tightly associated with heterochromatin in wt cells, enrichment of the TS domain at chromocenters was lost and the signal was diffusely distributed in the nucleus in UHRF1 depleted cells (*Uhrf1*^{-/-}) (Fig. 4C). A similar delocalization was found in UHRF1 binding deficient TS

domain mutants (Fig. 4D, Supplementary Material, Fig. S5). With half nucleus FRAP we investigated the mobility of TS domain fusions in MEFs and found that HSNIE mutations caused fast protein mobility indicating chromatin binding defects (Fig. 4E). Notably, the interaction of DNMT1 with the replication protein PCNA was not altered by the HSNIE mutations arguing against severe misfolding of the mutant proteins (Supplementary

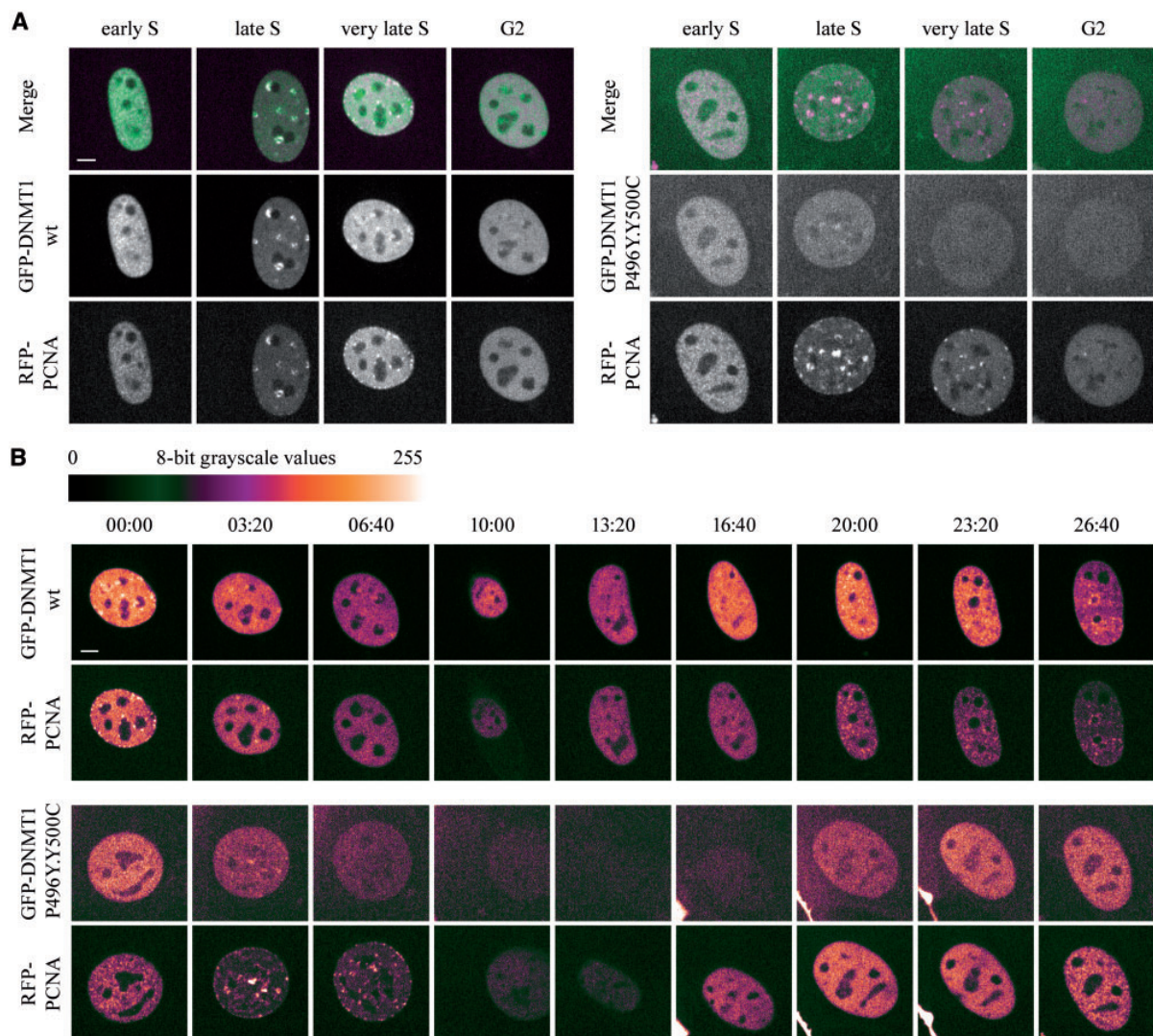


Figure 5. The destabilization of the DNMT1 HSNIE associated mutations is cell cycle dependent. (A and B) Spinning disk confocal midsections of MEF cells transiently coexpressing GFP-DNMT1 wt or P496Y.Y500C mutant and RFP-PCNA as a cell cycle marker. (A) Selected frames from live cell series are shown of GFP-DNMT1 wt (left panel) and GFP-DNMT1 P496Y.Y500C (right panel). Cells were tracked starting from early S until G2 phase. In the merged image, RFP-PCNA is depicted in magenta. (B) Live cell series of MEF cells shown in (A) transiently coexpressing GFP-DNMT1 wt (upper panel) or P496Y.Y500C (lower panel). Starting from very late S phase (wt) or mid S phase (P496Y.Y500C) images were taken every 200 min. White represents the highest and black the lowest intensity. Scale bar 5 μ m.

Material, Fig. S2). In conclusion, our findings suggest that the TS domain is crucial for mediating the interaction of DNMT1 with UHRF1 and that this interaction is affected by the HSNIE associated mutations located in this region.

The destabilization of HSNIE associated GFP-DNMT1 TS mutants is cell cycle-dependent

As the TS domain is essential for late S phase-specific localization and mobility of DNMT1 (42), we tested the effect of HSNIE mutations on subcellular localization of GFP-DNMT1 on a single cell level throughout the cell cycle. We imaged living cells coexpressing GFP-DNMT1 wt or GFP-DNMT1 P496Y.Y500C with RFP-PCNA as a cell cycle marker. While the localization in early to mid S phase was comparable to wt, the GFP signal of mutant

DNMT1 dropped when cells entered late S phase. In addition, the double point mutant displayed only weak late S phase-specific association with heterochromatin and did not display the characteristic, prolonged TS domain mediated heterochromatin association in G2 (Fig. 5A). Interestingly, the signal of GFP-DNMT1 P496Y.Y500C recovered in G1 phase indicating a cell cycle dependency of this defect (Fig. 5B). To investigate a potential stability defect of mutant DNMT1, we performed cycloheximide experiments, which showed that GFP-DNMT1 abundance and stability was reduced by the HSNIE mutations (P496Y, Y500C or P496Y.Y500C) (Supplementary Material, Fig. S6). These results provide evidence for a cell cycle-dependent destabilization of GFP-DNMT1 harboring HSNIE mutations beginning in late S phase that might be caused by insufficient targeting to late replicating heterochromatin.

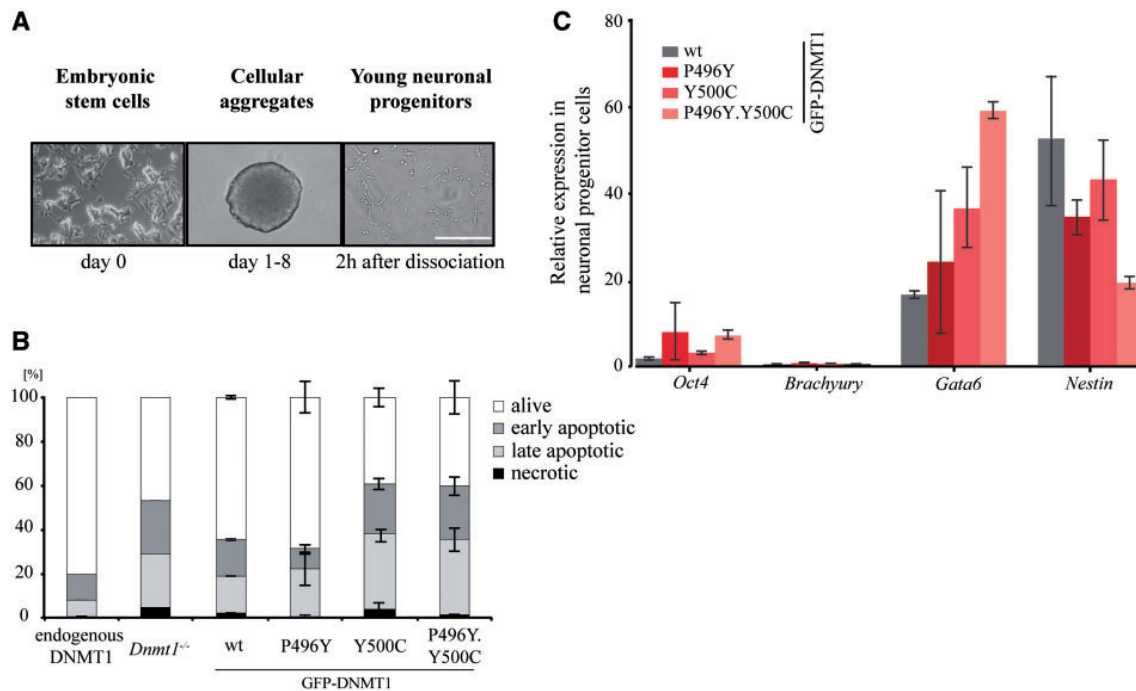


Figure 6. HSNIE associated point mutations in the TS domain show different survival rate and differentiation potential during neuronal progenitor differentiation. (A) Mouse embryonic stem cells are differentiated into young neuronal progenitor cells (NPCs). Scale bar: 200 μ m. (B) Flow cytometric analysis of alive, early apoptotic, late apoptotic and necrotic cells. Two hours after dissociation of cellular aggregates, young neuronal progenitors were analyzed using Annexin V and propidium iodide staining. Bar graphs represent mean values \pm standard deviation (SD) from two to three biological replicates. (C) RNA expression profiles of NPCs in biological duplicates (wt) or triplicates (P496Y, Y500C, P496Y.Y500C) of pluripotency factor *Oct4*, lineage specific markers *Brachyury*, *Gata6* and *Nestin* for mesodermal, endodermal and ectodermal lineage, respectively. All ddCt values are normalized to wt *Oct4* expression.

HSNIE associated point mutations affect neuronal progenitor differentiation

As HSNIE associated mutations in DNMT1 lead to a neuronal restricted phenotype in HSNIE patients (31), we investigated the potential of cells expressing DNMT1 with HSNIE associated mutations to differentiate into the neuronal lineage. During neuronal progenitor cell (NPC) differentiation (Fig. 6A), we examined cell viability of wt, *Dnmt1*^{-/-} and different ESC lines carrying HSNIE associated mutations (Supplementary Material, Fig. S7). In the pluripotent state and on day 8 of differentiation (Supplementary Material, Fig. S8), wt, *Dnmt1*^{-/-} and mutant cell lines exhibited a similar viability degree. Strikingly, young neuronal progenitors generated from *Dnmt1*^{-/-} cells, as well as stably expressing GFP-DNMT1 Y500C and double mutants P496Y.Y500C did not fully develop into differentiated NPCs, but instead were prone to undergo apoptosis 2 h after dissociation from cellular aggregates (Fig. 6B). Furthermore, we compared mRNA expression levels in NPCs of pluripotency gene *Oct4* and *Brachyury*, *Gata6* and *Nestin* for early mesodermal, primitive endodermal and ectodermal lineage, respectively. GFP-DNMT1 Y500C and double mutant P496Y.Y500C NPCs had lower expression levels of developmental marker *Nestin*, a marker for neuronal progenitor differentiation and elevated expression levels of pluripotency marker *Oct4* indicating an uncompleted differentiation in NPCs (Fig. 6C). The results highlight the importance of proper DNMT1 function during neurogenesis and indicate that the limited differentiation potential of DNMT1 deficient cells cannot be rescued by DNMT1 harboring HSNIE associated mutations.

Discussion

For long, DNA methylation has been regarded as a stable epigenetic mark set by the *de novo* DNA methyltransferases DNMT3A and DNMT3B during development and maintained after each round of DNA replication by DNMT1 (1–5). This simple view of DNA methylation, however, cannot explain why DNMT1 is expressed and required in postmitotic cells (27,28). Despite or because of the central and ubiquitous role in DNA methylation only very few human diseases were linked to heterozygous DNMT1 mutations that are mostly restricted to the TS domain within the regulatory NTD of DNMT1 (43). So far, two neurodegenerative diseases HSNIE and ADCA-DN have been linked to mutations in the TS domain of DNMT1 (31,32) but the underlying molecular mechanism remained elusive.

To investigate how simple amino acid exchanges outside the catalytic domain of DNMT1 may lead to global hypomethylation and neurodegenerative diseases, we transferred these mutations to a controlled ESC differentiation system and assayed for cellular defects. As UHRF1 plays a central role in the regulation of DNMT1, we first investigated whether the HSNIE mutations affect this interaction. Our coimmunoprecipitation experiments and deletion analyses show an interaction of the TS domain with the SRA domain of UHRF1, which is consistent with previous yeast two-hybrid screens (44). Interestingly, the HSNIE mutations are located within this part of the TS domain and indeed weaken this protein–protein interaction. We could demonstrate that mutated DNMT1 showed weaker association with heterochromatin in late S phase and failed to maintain DNA methylation in ESC, which is consistent with the previous

observation that UHRF1 is required for recruitment of DNMT1 (10). The interaction of UHRF1 with DNMT1 was also shown to be essential for allosteric activation of the enzyme to enable binding of substrate DNA to the CTD (23,24,45–47). The weakened interaction could explain the reduced activity of mutated DNMT1 pull-downs in vitro (31) and fits well with our observation that DNMT1 harboring HSANIE mutations are impaired in catalytic complex formation as measured with our trapping assay in living cells. In addition to defects in enzyme activation and targeting, we show that HSANIE mutations also cause proteasomal degradation of DNMT1 in late S and G2 phase, likely as a result of failed heterochromatin binding, which is consistent with previous studies (31). Similar protein destabilization was shown for chromatin unbound DNMT3A and DNMT3B (48,49).

Besides harboring a binding site for UHRF1, the TS domain of DNMT1 also contains a UIM that recognizes H3K18 ubiquitinated by UHRF1 (21). Although this UIM is well separated in primary and tertiary structure from the sites mutated in HSANIE patients, allosteric interference with H3K18ub binding cannot be ruled out. In addition to UHRF1 and histone binding, DNMT1 is subjected to further protein interactions and multiple post-translational modifications, which likely contribute to proper regulation of DNA methylation (50). Several of these interactions may not be absolutely required but may enhance local or global efficiency and thereby contribute to the fine tuning DNMT1 activity as we found for the interaction with PCNA, which 2-fold enhances DNMT1 efficiency in living cells (51,52). While it is unclear which of these interactions are affected by the HSANIE mutations in different cell types, we could clearly show that ESCs carrying these mutations are impaired in neuronal differentiation and prone to apoptosis.

Curiously, although HSANIE TS domain mutations are present in all patient tissues, they mostly affect the neuronal lineage causing both central and peripheral neurodegeneration. Although it is not clear yet how dynamic the changes in DNA methylation could be in postmitotic neurons, the turnover and change of DNA methylation has emerged as one possible modulator of neuronal plasticity in response to external or internal stimuli (53–55). This fits with the observation that neuronal tissues are characterized by strong expression of ten-eleven translocation (TET) genes and high hmC levels (56–59). Therefore, DNA modification might be more dynamic in postmitotic neurons than previously thought so that even small changes in protein–protein interactions and activity might unbalance the equilibrium of DNA modifications. But it remains unclear why neuronal tissues are specifically affected by these DNMT1 mutations in HSANIE patients.

In summary, we show that HSANIE mutations affect DNMT1 interaction with the essential cofactor UHRF1, cause cell cycle dependent degradation of DNMT1 and impair neuronal differentiation. These data add to our understanding of the role and regulation of DNMT1 during differentiation and help to understand DNMT1 dysfunction and hypomethylation in the pathogenesis of this neurodegenerative disease.

Materials and Methods

Mammalian expression constructs and antibodies

Mouse fusion constructs were generated using enhanced green fluorescent protein (GFP), monomeric red fluorescent protein (RFP) or monomeric cherry (Ch). The expression constructs for RFP-DNMT1, GFP-DNMT1 wt, GFP-DNMT1Δ356–404, GFP-

DNMT1Δ458–500, GFP-NTD, Ch-TS, GFP-DNMT1 (1–308), GFP-TS, GFP-DNMT1 (629–1110) and GFP-CTD have been described previously (9,21,26,40,60,61). GFP-DNMT1 deletion and HSANIE point mutants as well as UHRF1-GFP deletion expression constructs were derived from the corresponding wt constructs by overlap extension PCR (62). The GFP-UHRF1, UHRF1-GFP, Ch-UHRF1 and GFP-UHRF1 single domain constructs have been reported before (17,51,63,64). GFP, RFP and RFP-PCNA have been reported before (9,25,65,66). All constructs were verified by DNA sequencing. The following monoclonal antibodies were used for immunoblotting: rat anti-RFP/Ch (5F8, Chromotek; (67), rat anti-GFP (3H9, Chromotek). In dependence on the expected intensity of the signals, secondary antibodies either conjugated to horseradish peroxidase (HRP) (anti-rat (Dianova)) or conjugated to fluorescent dyes (anti-rat Alexa Fluor 488, Alexa Fluor 594 or Alexa Fluor 647 (Invitrogen)) were applied. For detection of HRP-conjugated antibodies an ECL Plus reagent (GE Healthcare, Thermo Scientific) was used.

Cell culture, transfection and immunofluorescence staining

HEK293T cells were cultured in Dulbecco's modified Eagle medium supplemented with 10% fetal bovine serum and 50 µg/ml gentamicin. MEF cells were cultured in Dulbecco's modified Eagle medium supplemented with 15% fetal bovine serum, non-essential amino acids, 2 mM L-glutamine, 0.1 mM β-mercaptoethanol (Gibco-BRL), 100 U/ml penicillin and 100 µg/ml streptomycin (PAA Laboratories GmbH). Mouse ESCs were cultured as published (51) with the exception that the medium was supplemented with 2i inhibitors (1 µM MEK inhibitor PD and 3 µM GSK-3 inhibitor CHIR) (2i, Axon Medchem) (63). To analyze the DNA methylation level in stably expressing UHRF1-GFP ESC lines, we cultured the cell lines in medium supplemented with 1,000 U/ml recombinant mouse leukemia inhibitory factor LIF (Millipore). Mouse J1 *Dnmt1*^{−/−} ESCs are homozygous for the c null allele and have been described before (65). Mouse E14 wt and *Uhrf1*^{−/−} cells (M. Muto and H. Koseki) as well as J1 triple knockout cells (Masaki Okano) have been reported previously (68). HEK293T cells were transfected with polyethylenimine (Sigma). Mouse ESCs and MEF cells were transfected with Lipofectamin (Invitrogen). Fixation, DAPI counterstaining and image acquisition cells was performed as described before (17).

Live cell microscopy, DNA methyltransferase trapping assay and fluorescence after photobleaching analysis

Live cell imaging and the trapping assay were performed as described previously (40,42). Briefly, mouse embryonic fibroblasts stably expressing RFP-PCNA were seeded on multi-well imaging slides (ibidi) and transfected using Lipofectamine 3000 (Life Technologies) following the manufacturer's protocol. Trapping assay was performed on an Ultraview-Vox spinning disk system (Perkin-Elmer) equipped with an EMCCD camera (Hamamatsu, Japan) a microirradiation system, and an environmental chamber kept at 37 °C with 5% CO₂, using a 63X 1.4 NA Plan-Apochromat oil-immersion objective (Zeiss). Imaging was performed with 488 nm and 561 nm solid-state laser diodes, using minimal gain and 2×2 camera binning for a final pixel size of 220 nm. Cells in late S-phase were visually identified for photobleaching experiments. For each photobleaching experiment, three pre-bleach images were acquired, before using the microirradiation system with the 488 nm laser to photobleach a

small area (<1 μm). After bleaching, images were acquired for both fluorophores every 10 s for up to 5 min.

Data extraction and FRAP normalization was performed in Fiji. Briefly, images were background subtracted and registered to correct for xy drift and cell motion artifacts. ROIs corresponding to the bleached area and to the entire nucleus were manually selected, and mean fluorescence intensities were extracted for each timepoint. From these raw intensities, two normalization steps were performed. First each data point was normalized to its corresponding average pre-bleach intensity, and then each data point from the bleached area was normalized to the total nuclear fluorescence at its corresponding timepoint to correct for acquisition photobleaching. Normalized recovery data was then imported into R (69).

Generation of stable ESC lines and DNA methylation analysis

Forty-eight hours after transfection with GFP tagged constructs, GFP positive ESCs were separated using a fluorescence activated cell sorting (FACS) Aria II instrument (Becton Dickinson) and the cells were subsequently grown in selective medium containing 10 $\mu\text{g}/\text{ml}$ blasticidin (GFP-DNMT1 cell lines) or normal medium (UHRF1-GFP cell lines). After expansion, cells were again FACS sorted one or two more times until at least 90% of the population was GFP positive. Furthermore, the GFP-DNMT1 cell lines were single cell sorted and clones with low expression levels were chosen for further analysis. The GFP-DNMT1 ESC line has been reported before (26). Genomic DNA isolation, bisulfite conversion, Primer sets and PCR conditions were described before (61,70). All PCR products were analyzed by pyrosequencing (Varionostic).

Neuronal progenitor cell (NPC) differentiation

The differentiation of pluripotent ESCs into NPCs was based on a protocol described before (71).

Analysis of cell viability

In order to analyze cell viability, we stained apoptotic cells with annexin V and necrotic cells with propidium iodide. To test the efficiency of the staining, apoptosis was induced by treatment of the cells with 5 μM staurosporine for 2 h ('apoptotic control') (Supplementary Material, Fig. S8). For the staining 200,000 cells were resuspended in 100 μl annexin binding buffer (100 mM HEPES pH 7.4, 140 mM NaCl, 2.5 mM CaCl_2). Cells were stained with 5 μl Alexa Fluor 350 conjugated annexin V (Thermo Fisher) for 15 min at room temperature protected from light. 400 μl annexin binding buffer were added and necrotic cells were stained by addition of 20 $\mu\text{g}/\text{ml}$ propidium iodide solution (Sigma Aldrich) shortly before analysis by flow cytometry using a FACS Aria II instrument (Becton Dickinson). Quantification of alive, necrotic, early and late apoptotic cells was performed using the single cell analysis software FlowJo.

RNA isolation, cDNA synthesis and real-time PCR

Total RNA was isolated from cells using a nucleospin triprep kit (Macherey-Nagel). 500 ng of total RNA was reverse transcribed with a high-capacity cDNA reverse transcription kit (Applied Biosystems) according to the manufacturer's instructions. Real-time PCR was conducted using LightCycler® 480 SYBR Green I

Master on a LightCycler® 480 Instrument II (Roche). PCR efficiency and primer pair specificity was examined using a standard curve of serially diluted cDNA and melting curve, respectively. After normalization to the transcript level of glyceraldehyde phosphate dehydrogenase, data were analyzed based on the $2^{-\Delta\Delta\text{CT}}$ method (72). A detailed list of primers used in the real-time PCR is shown in Supplementary Material, Table S1.

Protein-protein interaction assay and coimmunoprecipitation

For protein-protein interaction assays and coimmunoprecipitation GFP and RFP or Ch fusion constructs were expressed in HEK293T cells and 2 days after transfection cells were harvested in ice cold PBS. Cell pellets from one to two 10 cm dishes were lysed in 200 μl lysis buffer (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.5 mM EDTA, 0.1 mM MgCl_2 , 0.1% NP-40, 1 \times protease inhibitor, 2 mM PMSF, 1 mg/ml DNaseI (AppliChem)) and a protein-protein interaction assay in GFP-multiTrap plates (Chromotek) was performed as described (41) with the following adaptations: GFP extracts were equalized to a concentration of 60 nM in immunoprecipitation buffer (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.5 mM EDTA) prior to one step purification in blocked (3% milk) GFP-multiTrap plates. After stringent washing (wash buffer; 20 mM Tris-HCl pH 7.5, 300 mM NaCl, 0.5 mM EDTA), purified GFP fusion proteins were incubated with crude protein extracts of RFP or Ch fusion proteins at a concentration of 1.1 to 2.1 μM diluted in IP buffer (excess of amount RFP or Ch fusion proteins in relation to GFP fusions: 18–35 times). Bound fractions were quantified by fluorescence intensity measurements with a Tecan Infinite M1000 plate reader (Tecan). For coimmunoprecipitation assays, the GFP and RFP or Ch fusion constructs were coexpressed in HEK293T cells, protein extracts were equalized and depending on the expression amounts of 5–30 pmol GFP-fusion protein were applied for the coimmunoprecipitation with the GFP-Trap (Chromotek). Note that the plasmid amount of GFP fusion construct and RFP fusion constructs used for transfection was adapted in a way to have at least a 3-fold excess of the molar RFP or Ch fusion protein amount in relation to GFP fusions. Bound fractions were firstly detected by fluorescence intensity measurements and secondly by immunoblotting using specific antibodies.

Statistical Analysis

Results were depicted as mean values \pm standard deviations (SDs) or as mean values \pm standard errors of the mean (SEM) from the number of biological replicates indicated in the corresponding figure legend. The difference between two mean values was analyzed by Student's t-test and was considered as statistically significant in case of $P < 0.05$ (*) and highly significant for $P < 0.001$ (**).

Sequence Alignments

Alignments were prepared using ClustalW2 (73) and ESPript (74).

Supplementary Material

Supplementary Material is available at HMG online.

Acknowledgements

We are grateful to the following colleagues for providing ESCs and somatic cells: Masahiro Muto and Haruhiko Koseki for mouse wt E14 and *Uhrf1*^{-/-} ESCs; En Li and Taiping Chen for mouse J1 wt and *Dnmt1*^{-/-} ESCs; Masaki Okano for J1 TKO ESCs; Thomas Jenuwein for wt MEF cells.

Conflict of Interest statement. None declared.

Funding

This work was supported by grants from the Deutsche Forschungsgemeinschaft [grant number SFB 646/B10, and SFB1064/A17 to H.L.]. M.S. and S.L. are fellows of the Integrated Research Training Group (IRTG) of the SFB1064. S.L. was funded by an award of Lehre@LMU of the Ludwig-Maximilians Universität Munich. P.W. was a fellow of the Graduate School Life Science Munich (LSM). K.S. was and J.R. is supported by the International Max Planck Research School for Molecular and Cellular Life Sciences (IMPRS-LS). J.R. is supported by a doctoral fellowship from Fonds de recherche du Québec - Santé. W.Q. was supported by the China Scholarship Council (CSC). Funding to pay the Open Access publication charges for this article was provided by the Deutsche Forschungsgemeinschaft [grant number SFB1064/A17 to H.L.].

References

- Bird, A. (2002) DNA methylation patterns and epigenetic memory. *Genes Dev.*, **16**, 6–21.
- Goll, M.G. and Bestor, T.H. (2005) Eukaryotic cytosine methyltransferases. *Annu. Rev. Biochem.*, **74**, 481–514.
- Cheng, X. and Blumenthal, R.M. (2008) Mammalian DNA methyltransferases: a structural perspective. *Structure*, **16**, 341–350.
- Law, J.A. and Jacobsen, S.E. (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.*, **11**, 204–220.
- Smith, Z.D. and Meissner, A. (2013) DNA methylation: roles in mammalian development. *Nat. Rev. Genet.*, **14**, 204–220.
- Margot, J.B., Aguirre-Arteta, A.M., Di Giacco, B.V., Pradhan, S., Roberts, R.J., Cardoso, M.C. and Leonhardt, H. (2000) Structure and function of the mouse DNA methyltransferase gene: *Dnmt1* shows a tripartite structure. *J. Mol. Biol.*, **297**, 293–300.
- Leonhardt, H., Page, A.W., Weier, H.U. and Bestor, T.H. (1992) A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. *Cell*, **71**, 865–873.
- Chuang, L.S., Ian, H.I., Koh, T.W., Ng, H.H., Xu, G. and Li, B.F. (1997) Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. *Science*, **277**, 1996–2000.
- Easwaran, H.P., Schermelleh, L., Leonhardt, H. and Cardoso, M.C. (2004) Replication-independent chromatin loading of *Dnmt1* during G2 and M phases. *EMBO Rep.*, **5**, 1181–1186.
- Bostick, M., Kim, J.K., Estève, P.O., Clark, A., Pradhan, S. and Jacobsen, S.E. (2007) UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science*, **317**, 1760–1764.
- Sharif, J., Muto, M., Takebayashi, S.I., Suetake, I., Iwamatsu, A., Endo, T.A., Shinga, J., Mizutani-Koseki, Y., Toyoda, T., Okamura, K., et al. (2007) The SRA protein Np95 mediates epigenetic inheritance by recruiting *Dnmt1* to methylated DNA. *Nature*, **450**, 908–912.
- Arita, K., Ariyoshi, M., Tochio, H., Nakamura, Y. and Shirakawa, M. (2008) Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. *Nature*, **455**, 818–821.
- Avvakumov, G.V., Walker, J.R., Xue, S., Li, Y., Duan, S., Bronner, C., Arrowsmith, C.H. and Dhe-Paganon, S. (2008) Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. *Nature*, **455**, 822–825.
- Qian, C., Li, S., Jakoncic, J., Zeng, L., Walsh, M.J. and Zhou, M.M. (2008) Structure and hemimethylated CpG binding of the SRA domain from human UHRF1. *J. Biol. Chem.*, **283**, 34490–34494.
- Citterio, E., Papait, R., Nicassio, F., Vecchi, M., Gomiero, P., Mantovani, R., Di Fiore, P.P. and Bonapace, I.M. (2004) Np95 is a histone-binding protein endowed with ubiquitin ligase activity. *Mol. Cell. Biol.*, **24**, 2526–2535.
- Karagianni, P., Amazit, L., Qin, J. and Wong, J. (2008) ICBP90, a novel methyl K9 H3 binding protein linking protein ubiquitination with heterochromatin formation. *Mol. Cell. Biol.*, **28**, 705–717.
- Rottach, A., Frauer, C., Pichler, G., Bonapace, I.M., Spada, F. and Leonhardt, H. (2010) The multi-domain protein Np95 connects DNA methylation and histone modification. *Nucleic Acids Res.*, **38**, 1796–1804.
- Cheng, J., Yang, Y., Fang, J., Xiao, J., Zhu, T., Chen, F., Wang, P., Li, Z., Yang, H. and Xu, Y. (2013) Structural insight into coordinated recognition of trimethylated histone H3 lysine 9 (H3K9me3) by the plant homeodomain (PHD) and tandem tudor domain (TTD) of UHRF1 (ubiquitin-like, containing PHD and RING finger domains, 1) protein. *J. Biol. Chem.*, **288**, 1329–1339.
- Liu, X., Gao, Q., Li, P., Zhao, Q., Zhang, J., Li, J., Koseki, H. and Wong, J. (2013) UHRF1 targets DNMT1 for DNA methylation through cooperative binding of hemi-methylated DNA and methylated H3K9. *Nat. Commun.*, **4**, 1563.
- Nishiyama, A., Yamaguchi, L., Sharif, J., Johmura, Y., Kawamura, T., Nakanishi, K., Shimamura, S., Arita, K., Kodama, T., Ishikawa, F., et al. (2013) Uhrf1-dependent H3K23 ubiquitylation couples maintenance DNA methylation and replication. *Nature*, **502**, 249–253.
- Qin, W., Wolf, P., Liu, N., Link, S., Smets, M., La Mastra, F., Forné, I., Pichler, G., Hörl, D., Fellingner, K., et al. (2015) DNA methylation requires a DNMT1 ubiquitin interacting motif (UIM) and histone ubiquitination. *Cell Res.*, **25**, 911–929.
- Song, J., Rechko, O., Bestor, T.H. and Patel, D.J. (2010) Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. *Science*, **331**, 1036–1040.
- Takeshita, K., Suetake, I., Yamashita, E., Suga, M., Narita, H., Nakagawa, A. and Tajima, S. (2011) Structural insight into maintenance methylation by mouse DNA methyltransferase 1 (*Dnmt1*). *Proc. Natl. Acad. Sci. USA*, **108**, 9055–9059.
- Bashtrykov, P., Jankevicius, G., Jurkowska, R.Z., Ragozin, S. and Jeltsch, A. (2014) The UHRF1 protein stimulates the activity and specificity of the maintenance DNA methyltransferase DNMT1 by an allosteric mechanism. *J. Biol. Chem.*, **289**, 4106–4115.
- Du, Z., Song, J., Wang, Y., Zhao, Y., Guda, K., Yang, S., Kao, H.Y., Xu, Y., Willis, J., Markowitz, S.D., et al. (2010) DNMT1 stability is regulated by proteins coordinating deubiquitination and acetylation-driven ubiquitination. *Sci. Signal.*, **3**, ra80.

26. Qin, W., Leonhardt, H. and Spada, F. (2011) Usp7 and Uhrf1 control ubiquitination and stability of the maintenance DNA methyltransferase Dnmt1. *J. Cell. Biochem.*, **112**, 439–444.
27. Goto, K., Numata, M., Komura, J.I., Ono, T., Bestor, T.H. and Kondo, H. (1994) Expression of DNA methyltransferase gene in mature and immature neurons as well as proliferating cells in mice. *Differentiation*, **56**, 39–44.
28. Inano, K., Suetake, I., Ueda, T., Miyake, Y., Nakamura, M., Okada, M. and Tajima, S. (2000) Maintenance-type DNA methyltransferase is highly expressed in post-mitotic neurons and localized in the cytoplasmic compartment. *J. Biochem.*, **128**, 315–321.
29. Zhao, C., Deng, W. and Gage, F.H. (2008) Mechanisms and functional implications of adult neurogenesis. *Cell*, **132**, 645–660.
30. Braun, S.M.G. and Jessberger, S. (2014) Adult neurogenesis and its role in neuropsychiatric disease, brain repair and normal brain function. *Neuropathol. Appl. Neurobiol.*, **40**, 3–12.
31. Klein, C.J., Botuyan, M.V., Wu, Y., Ward, C.J., Nicholson, G.A., Hammans, S., Hojo, K., Yamanishi, H., Karpf, A.R., Wallace, D.C., et al. (2011) Mutations in DNMT1 cause hereditary sensory neuropathy with dementia and hearing loss. *Nat. Genet.*, **43**, 595–600.
32. Winkelmann, J., Lin, L., Schormair, B., Kornum, B.R., Faraco, J., Plazzi, G., Melberg, A., Cornelio, F., Urban, A.E., Pizza, F., et al. (2012) Mutations in DNMT1 cause autosomal dominant cerebellar ataxia, deafness and narcolepsy. *Hum. Mol. Genet.*, **21**, 2205–2210.
33. David, G., Gosal, D., Ealing, J. and Mignot, E. (2013) A mutation in the DNMT1 gene causing autosomal dominant ataxia with deafness and cataplexy. *J. Neurol. Neurosurg. Psychiatry*, **84**, e2.44–e2.
34. Pedrosa, J.L., Povoas Barsottini, O.G., Lin, L., Melberg, A., Oliveira, A.S.B. and Mignot, E. (2013) A novel de novo exon 21 DNMT1 mutation causes cerebellar ataxia, deafness, and narcolepsy in a Brazilian patient. *Sleep*, **36**, 1257–1259.
35. Yuan, J., Higuchi, Y., Nagado, T., Nozuma, S., Nakamura, T., Matsuura, E., Hashiguchi, A., Sakiyama, Y., Yoshimura, A. and Takashima, H. (2013) Novel mutation in the replication focus targeting sequence domain of DNMT1 causes hereditary sensory and autonomic neuropathy 1E. *J. Peripher. Nerv. Syst.*, **18**, 89–93.
36. Moghadam, K.K., Pizza, F., La Morgia, C., Franceschini, C., Tonon, C., Lodi, R., Barboni, P., Seri, M., Ferrari, S., Liguori, R., et al. (2014) Narcolepsy is a common phenotype in HSN1 IE and ADCA-DN. *Brain*, **137**, 1643–1655.
37. Sun, Z., Wu, Y., Ordog, T., Baheti, S., Nie, J., Duan, X., Hojo, K., Kocher, J.P., Dyck, P.J. and Klein, C.J. (2014) Aberrant signature methylome by DNMT1 hot spot mutation in hereditary sensory and autonomic neuropathy 1E. *Epigenetics*, **9**, 1184–1193.
38. Baets, J., Duan, X., Wu, Y., Smith, G., Seeley, W.W., Mademan, I., McGrath, N.M., Beadell, N.C., Khoury, J., Botuyan, M.V., et al. (2015) Defects of mutant DNMT1 are linked to a spectrum of neurological disorders. *Brain*, **138**, 845–861.
39. Klein, C.J., Bird, T., Ertekin-Taner, N., Lincoln, S., Hjorth, R., Wu, Y., Kwok, J., Mer, G., Dyck, P.J. and Nicholson, G.A. (2013) DNMT1 mutation hot spot causes varied phenotypes of HSN1 with dementia and hearing loss. *Neurology*, **80**, 824–828.
40. Schermelleh, L., Spada, F., Easwaran, H.P., Zolghadr, K., Margot, J.B., Cardoso, M.C. and Leonhardt, H. (2005) Trapped in action: direct visualization of DNA methyltransferase activity in living cells. *Nat. Methods*, **2**, 751–756.
41. Pichler, G., Jack, A., Wolf, P. and Hake, S.B. (2012) Versatile toolbox for high throughput biochemical and functional studies with fluorescent fusion proteins. *PLoS One*, **7**, e36967.
42. Schneider, K., Fuchs, C., Dobay, A., Rottach, A., Qin, W., Wolf, P., Álvarez-Castro, J.M., Nalaskowski, M.M., Kremmer, E., Schmid, V., et al. (2013) Dissection of cell cycle-dependent dynamics of Dnmt1 by FRAP and diffusion-coupled modeling. *Nucleic Acids Res.*, **41**, 4860–4876.
43. Bestor, T.H., Edwards, J.R. and Boulard, M. (2015) Notes on the role of dynamic DNA methylation in mammalian development. *Proc. Natl. Acad. Sci. USA*, **112**, 6796–6799.
44. Achour, M., Jacq, X., Rondé, P., Alhosin, M., Charlot, C., Chataigneau, T., Jeanblanc, M., Macaluso, M., Giordano, A., Hughes, A.D., et al. (2008) The interaction of the SRA domain of ICBP90 with a novel domain of DNMT1 is involved in the regulation of VEGF gene expression. *Oncogene*, **27**, 2187–2197.
45. Syeda, F., Fagan, R.L., Wean, M., Avvakumov, G.V., Walker, J.R., Xue, S., Dhe-Paganon, S. and Brenner, C. (2011) The replication focus targeting sequence (RFTS) domain is a DNA-competitive inhibitor of Dnmt1. *J. Biol. Chem.*, **286**, 15344–15351.
46. Bashtrykov, P., Rajavelu, A., Hackner, B., Ragozin, S., Carell, T. and Jeltsch, A. (2014) Targeted mutagenesis results in an activation of DNA methyltransferase 1 and confirms an autoinhibitory role of its RFTS domain. *Chembiochem*, **15**, 743–748.
47. Berkuyrek, A.C., Suetake, I., Arita, K., Takeshita, K., Nakagawa, A., Shirakawa, M. and Tajima, S. (2014) The DNA methyltransferase Dnmt1 directly interacts with the SET and RING finger-associated (SRA) domain of the multifunctional protein Uhrf1 to facilitate accession of the catalytic center to hemi-methylated DNA. *J. Biol. Chem.*, **289**, 379–386.
48. Jeong, S., Liang, G., Sharma, S., Lin, J.C., Choi, S.H., Han, H., Yoo, C.B., Egger, G., Yang, A.S. and Jones, P.A. (2009) Selective anchoring of DNA methyltransferases 3A and 3B to nucleosomes containing methylated DNA. *Mol. Cell. Biol.*, **29**, 5366–5376.
49. Sharma, S., De Carvalho, D.D., Jeong, S., Jones, P.A. and Liang, G. (2011) Nucleosomes containing methylated DNA stabilize DNA methyltransferases 3A/3B and ensure faithful epigenetic inheritance. *PLoS Genet.*, **7**, e1001286.
50. Qin, W., Leonhardt, H. and Pichler, G. (2011) Regulation of DNA methyltransferase 1 by interactions and modifications. *Nucleus*, **2**, 392–402.
51. Schermelleh, L., Haemmer, A., Spada, F., Rösing, N., Meilinger, D., Rothbauer, U., Cardoso, M.C. and Leonhardt, H. (2007) Dynamics of Dnmt1 interaction with the replication machinery and its role in postreplicative maintenance of DNA methylation. *Nucleic Acids Res.*, **35**, 4301–4312.
52. Spada, F., Haemmer, A., Kuch, D., Rothbauer, U., Schermelleh, L., Kremmer, E., Carell, T., Längst, G. and Leonhardt, H. (2007) DNMT1 but not its interaction with the replication machinery is required for maintenance of DNA methylation in human cells. *J. Cell Biol.*, **176**, 565–571.
53. Fan, G., Beard, C., Chen, R.Z., Csankovszki, G., Sun, Y., Siniaia, M., Biniszkiwicz, D., Bates, B., Lee, P.P., Kuhn, R., et al. (2001) DNA hypomethylation perturbs the function and survival of CNS neurons in postnatal animals. *J. Neurosci.*, **21**, 788–797.
54. Borrelli, E., Nestler, E.J., Allis, C.D. and Sassone-Corsi, P. (2008) Decoding the epigenetic language of neuronal plasticity. *Neuron*, **60**, 961–974.

55. Yu, N.K., Baek, S.H. and Kaang, B.K. (2011) DNA methylation-mediated control of learning and memory. *Mol. Brain*, **4**, 5.
56. Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., et al. (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*, **324**, 930–935.
57. Globisch, D., Münzel, M., Müller, M., Michalakakis, S., Wagner, M., Koch, S., Brückl, T., Biel, M. and Carell, T. (2010) Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS One*, **5**, e15367.
58. Szwagierczak, A., Bultmann, S., Schmidt, C.S., Spada, F. and Leonhardt, H. (2010) Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Res.*, **38**, e181.
59. Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C. and Zhang, Y. (2011) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science*, **333**, 1300–1303.
60. Fellinger, K., Rothbauer, U., Felle, M., Längst, G. and Leonhardt, H. (2009) Dimerization of DNA methyltransferase 1 is mediated by its regulatory domain. *J. Cell. Biochem.*, **106**, 521–528.
61. Frauer, C., Rottach, A., Meilinger, D., Bultmann, S., Fellinger, K., Hasenöder, S., Wang, M., Qin, W., Söding, J., Spada, F., et al. (2011) Different binding properties and function of CXXC zinc finger domains in Dnmt1 and Tet1. *PLoS One*, **6**, e16627.
62. Ho, S.N., Hunt, H.D., Horton, R.M., Pullen, J.K. and Pease, L.R. (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene*, **77**, 51–59.
63. Ying, Q.L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P. and Smith, A. (2008) The ground state of embryonic stem cell self-renewal. *Nature*, **453**, 519–523.
64. De Vos, M., El Ramy, R., Delphine, Q., Patricia, W., Fabio, S., Najat, M., Federica, B., Valérie, S., Heinrich, L., Bonapace, I.M., et al. (2014) Poly(ADP-ribose) Polymerase 1 (PARP1) Associates with E3 ubiquitin-protein ligase UHRF1 and modulates UHRF1 biological functions. *J. Biol. Chem.*, **289**, 16223–16238.
65. Lei, H., Oh, S.P., Okano, M., Jüttermann, R., Goss, K.A., Jaenisch, R. and Li, E. (1996) De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. *Development*, **122**, 3195–3205.
66. Schneider, C.A., Rasband, W.S. and Eliceiri, K.W. (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods*, **9**, 671–675.
67. Rottach, A., Kremmer, E., Nowak, D., Leonhardt, H. and Cardoso, M.C. (2008) Generation and characterization of a rat monoclonal antibody specific for multiple red fluorescent proteins. *Hybridoma*, **27**, 337–343.
68. Meilinger, D., Fellinger, K., Bultmann, S., Rothbauer, U., Bonapace, I.M., Klinkert, W.E.F., Spada, F. and Leonhardt, H. (2009) Np95 interacts with de novo DNA methyltransferases, Dnmt3a and Dnmt3b, and mediates epigenetic silencing of the viral CMV promoter in embryonic stem cells. *EMBO Rep.*, **10**, 1259–1264.
69. R Core Team (2016) R: A Language and Environment for Statistical Computing. The R Foundation for Statistical Computing, Vienna, Austria.
70. Tucker, K.L., Beard, C., Dausmann, J., Jackson-Grusby, L., Laird, P.W., Lei, H., Li, E. and Jaenisch, R. (1996) Germ-line passage is required for establishment of methylation and expression patterns of imprinted but not of nonimprinted genes. *Genes Dev.*, **10**, 1008–1020.
71. Bibel, M., Richter, J., Lacroix, E. and Barde, Y.A. (2007) Generation of a defined and uniform population of CNS progenitors and neurons from mouse embryonic stem cells. *Nat. Protoc.*, **2**, 1034–1043.
72. Livak, K.J. and Schmittgen, T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2- $\Delta\Delta$ CT method. *Methods*, **25**, 402–408.
73. Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.*, **7**, 539.
74. Robert, X. and Gouet, P. (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res.*, **42**, W320–W324.

Supplementary Data

DNMT1 Mutations found in HSNIE patients affect interaction with UHRF1 and neuronal differentiation

Martha Smets^{1#}, Stephanie Link^{1,2#}, Patricia Wolf^{1,3}, Katrin Schneider^{1,4}, Veronica Solis^{1,5}, Joel Ryan¹, Daniela Meilinger¹, Weihua Qin¹ and Heinrich Leonhardt^{1*}

¹Department of Biology II and Center for Integrated Protein Science Munich (CIPSM), Ludwig-Maximilians-Universität München, Großhaderner Str. 2, 82152 Planegg-Martinsried, Germany

Current addresses:

²BioMedical Center (BMC), Department of Molecular Biology, Ludwig-Maximilians-Universität München, Großhaderner Str. 9, 82152 Planegg-Martinsried, Germany

³Microcoat Biotechnologie GmbH, Am Neuland 3, 82347 Bernried am Starnberger See, Germany

⁴Definiens AG, Bernhard-Wicki-Str. 5, 80636 München, Germany

⁵Max Planck Institute of Neurobiology, Am Klopferspitz 18, 82152 Martinsried, Germany

[#]The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors

*Correspondence: Tel: +49 89 2180 74232; Fax: +49 89 2180 74236, h.leonhardt@lmu.de (H.L.)

Supplementary materials and methods

Mammalian expression constructs and antibodies

The construct GFP-DNMT1 Δ PBD has been reported previously (1). For immunofluorescent stainings of endogenous proteins, the monoclonal rat anti-DNMT1 antibody 5A10 (2) and a polyclonal rabbit anti-UHRF1 antibody were used (3). The following monoclonal antibodies were used for immunoblotting: rabbit anti-H3 (Abcam), rabbit anti-UHRF1 (3), and mouse anti β -Actin (Sigma) and rat anti-DNMT1 14F6 (4). As secondary antibodies an anti-rabbit Alexa Fluor 488 or anti-rat Alexa Fluor 594 were applied (Invitrogen). In dependence on the expected intensity of the signals, secondary antibodies either conjugated to horseradish peroxidase (HRP) (anti-rat (Dianova)) or conjugated to fluorescent dyes (anti-rat Alexa Fluor 488 or Alexa Fluor 647 (Invitrogen)) were applied. For detection of HRP-conjugated antibodies an ECL Plus reagent (GE Healthcare, Thermo Scientific) was used.

Cycloheximide assay and preparation of protein extracts

Mouse ESCs were plated at equal densities three hours before cycloheximide (CHX) (0.03 mg/ml, Sigma-Aldrich) and MG132 (5 μ M, Cell Signaling) treatment. DMSO served as mock control. Cell pellets were lysed in lysis buffer (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 2 mM MgCl₂, 0.5% NP40, 0.3 U/ μ l Benzonase (Sigma-Aldrich), 2 mM PMSF and protease inhibitor mix M (SERVA)) and lysates were equalized to the overall protein concentration using the Pierce™ 660nm Protein Assay Reagent (ThermoFisher). Protein levels were detected by immunoblotting with specific antibodies against DNMT1 (14F6) (4), UHRF1 (3), β -Actin (Sigma) and H3 (Abcam) and quantified using the ImageJ gel analysis tool.

Supplementary references

1. Easwaran, H.P., Schermelleh, L., Leonhardt, H. and Cardoso, M.C. (2004) Replication-independent chromatin loading of Dnmt1 during G2 and M phases. *EMBO Rep.*, **5**, 1181–1186.
2. Schneider, K., Fuchs, C., Dobay, A., Rottach, A., Qin, W., Wolf, P., Álvarez-Castro, J.M., Nalaskowski, M.M., Kremmer, E., Schmid, V., *et al.* (2013) Dissection of cell cycle-dependent dynamics of Dnmt1 by FRAP and diffusion-coupled modeling. *Nucleic Acids Res.*, **41**, 4860–4876.
3. Citterio, E., Papait, R., Nicassio, F., Vecchi, M., Gomiero, P., Mantovani, R., Di Fiore, P.P. and Bonapace, I.M. (2004) Np95 is a histone-binding protein endowed with ubiquitin ligase activity. *Mol. Cell. Biol.*, **24**, 2526–2535.
4. Mulholland, C.B., Smets, M., Schmidtman, E., Leidescher, S., Markaki, Y., Hofweber, M., Qin, W., Manzo, M., Kremmer, E., Thanisch, K., *et al.* (2015) A modular open platform for systematic functional studies under physiological conditions. *Nucleic Acids Res.*, **43**, e112.

Supplementary figures and legends

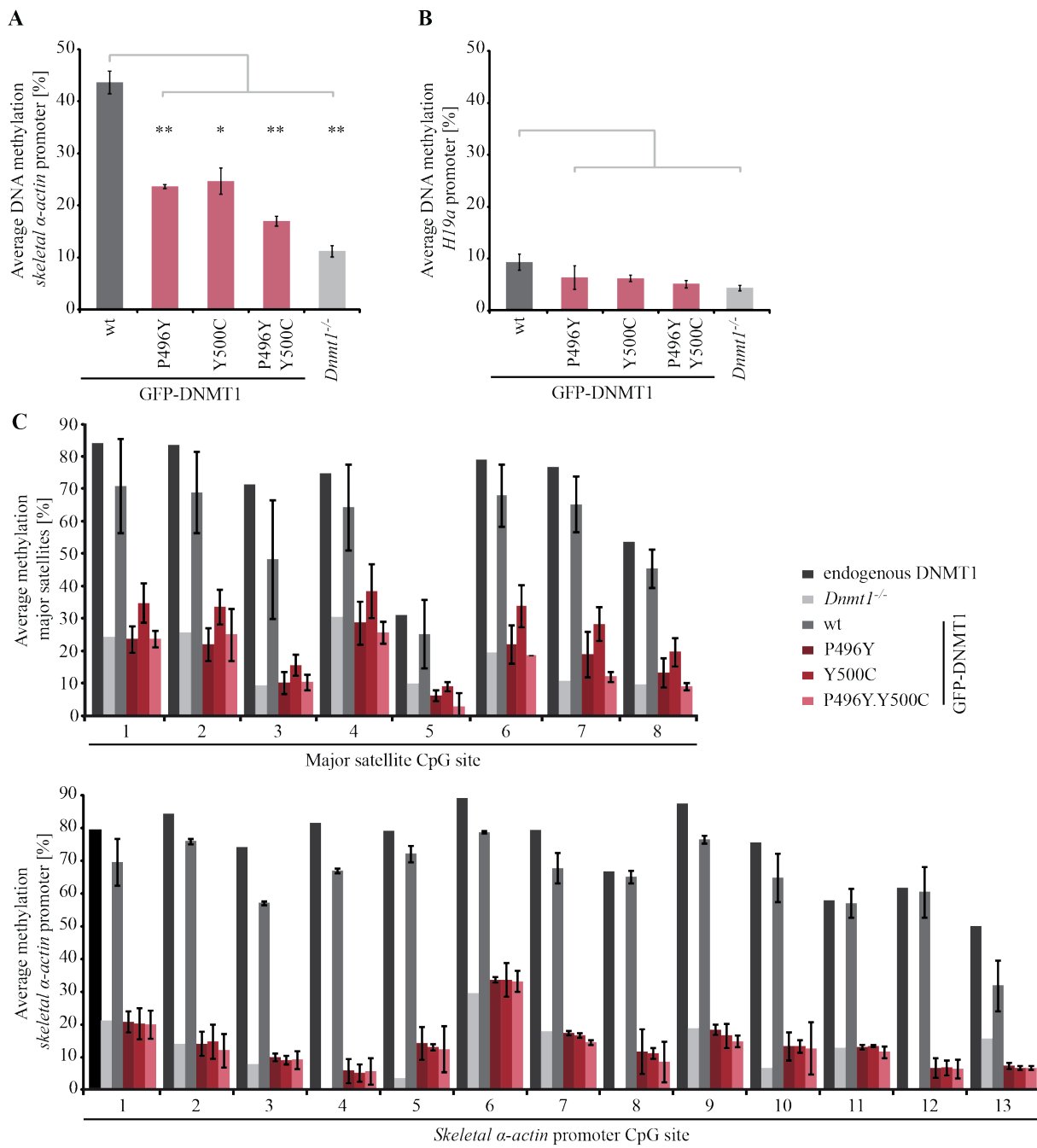


Figure S1. HSNIE associated DNMT1 TS mutants cannot restore the local methylation level at the *skeletal α -actin* promoter. (A and B) Site-specific DNA methylation in mouse *Dnmt1*^{-/-} ESCs after reintroducing GFP-DNMT1 wt or P496Y and Y500C as well as P496Y.Y500C of (A) the *skeletal α -actin* and (B) the *H19a*

promoter. Bar graphs represent mean values \pm s.d. from three biological replicates (average from 13 or 6 CpG sites, respectively) and two-sample t-tests were performed that assume equal variances. Statistical significance compared to the methylation level of GFP-DNMT1 wt is indicated: * $P < 0.02$, ** $P < 0.0002$. (C) CpG methylation levels at the major satellite repeats and the *skeletal α -actin* promoter of mouse *Dnmt1*^{-/-} ESCs stably expressing GFP-DNMT1 wt or HSNIE associated point mutants were analyzed. Shown are mean values from two different single cell clones, respectively.

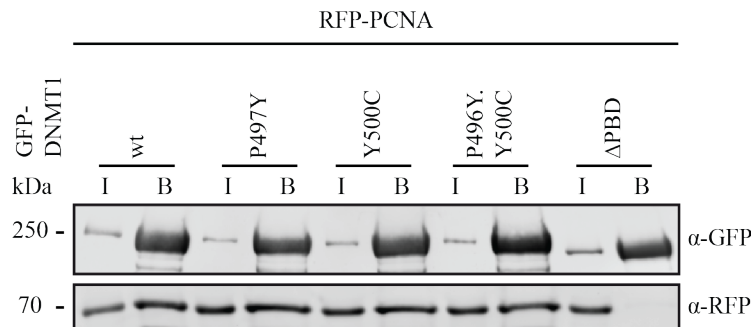


Figure S2. HSNIE associated point mutations in the DNMT1 TS domain do not affect the interaction with the replication protein PCNA. Coimmunoprecipitation of GFP-DNMT1 wt, P496Y, Y500C or P496Y.Y500C and RFP-PCNA. GFP and RFP fusion constructs were coexpressed in HEK293T cells and cell lysates were used for immunoprecipitation with the GFP-Trap. Bound fractions were analyzed by immunoblotting with anti-GFP and anti-RFP antibodies. GFP-DNMT1 lacking the PCNA-binding domain (PBD) served as a negative control. I = Input, B = Bound.

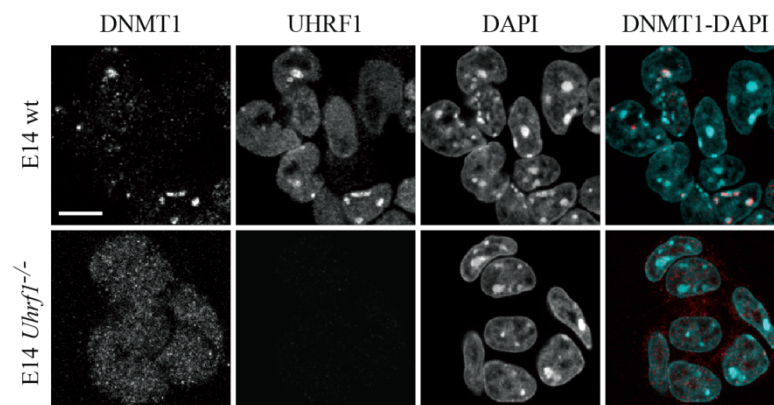


Figure S3. DNMT1 shows a diffuse nuclear pattern in *Uhrf1*^{-/-} ESCs. Confocal midsections of fixed mouse E14 wt and *Uhrf1*^{-/-} ESCs. Endogenous DNMT1 and UHRF1 were immunostained with specific antibodies and

chromatin was counterstained with DAPI. In the merged image, DAPI is depicted in magenta and DNMT1 in green. Scale bar 10 μ m.

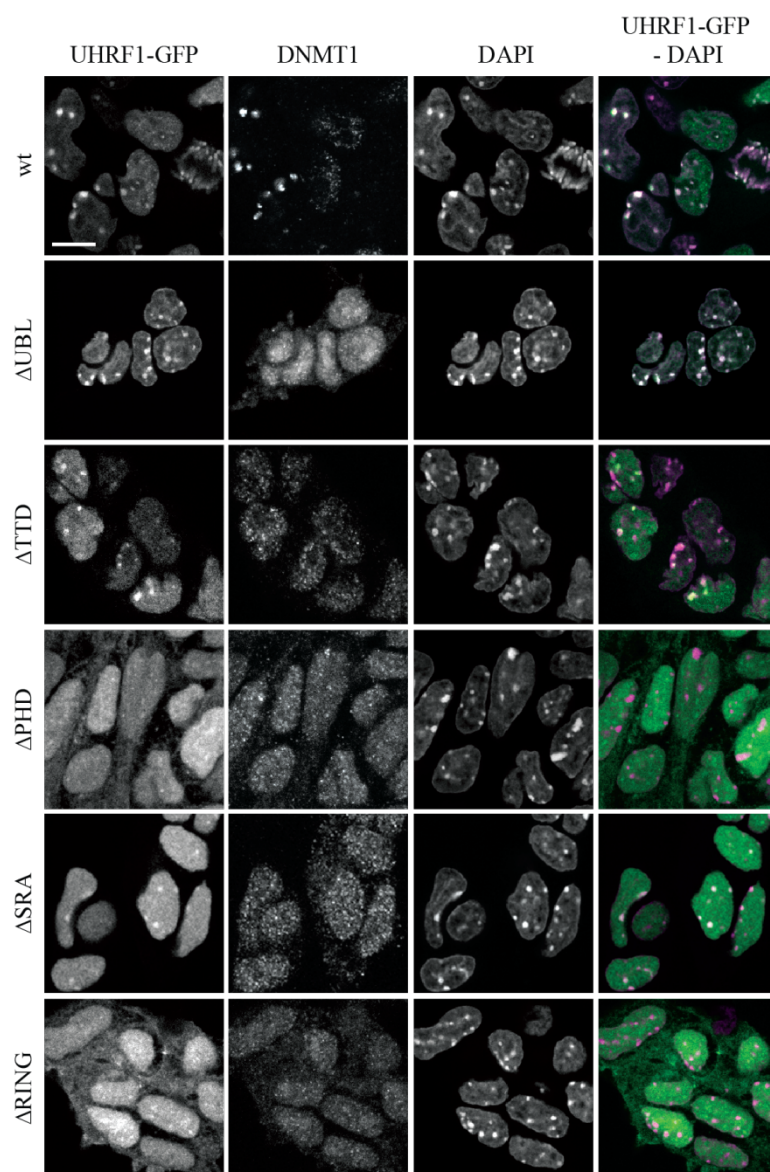


Figure S4. DNMT1 shows a diffuse pattern in *Uhrf1*^{-/-} ESCs stably expressing GFP-UHRF1 single domain deletion mutants. Confocal midsections of fixed mouse E14 *Uhrf1*^{-/-} ESCs stably expressing UHRF1-GFP wt or single domain deletion mutant proteins. Endogenous DNMT1 was immunostained with a specific antibody and chromatin was counterstained with DAPI. In the merged image, DAPI is depicted in magenta. Scale bar 10 μ m.

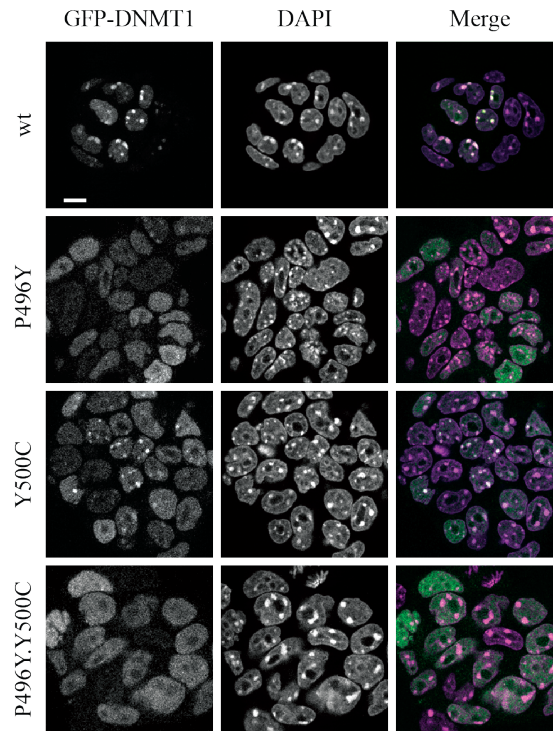


Figure S5. GFP-DNMT1 TS mutations lead to decreased association with heterochromatin. Localization of GFP-DNMT1 wt and HSNIE associated mutants P496Y, Y500C and P496Y.Y500C stably expressed in J1 *Dnmt1*^{-/-} ESCs. DAPI was used for chromatin counterstaining. In the merged image, DAPI is depicted in magenta. Scale bar 5 μ m.

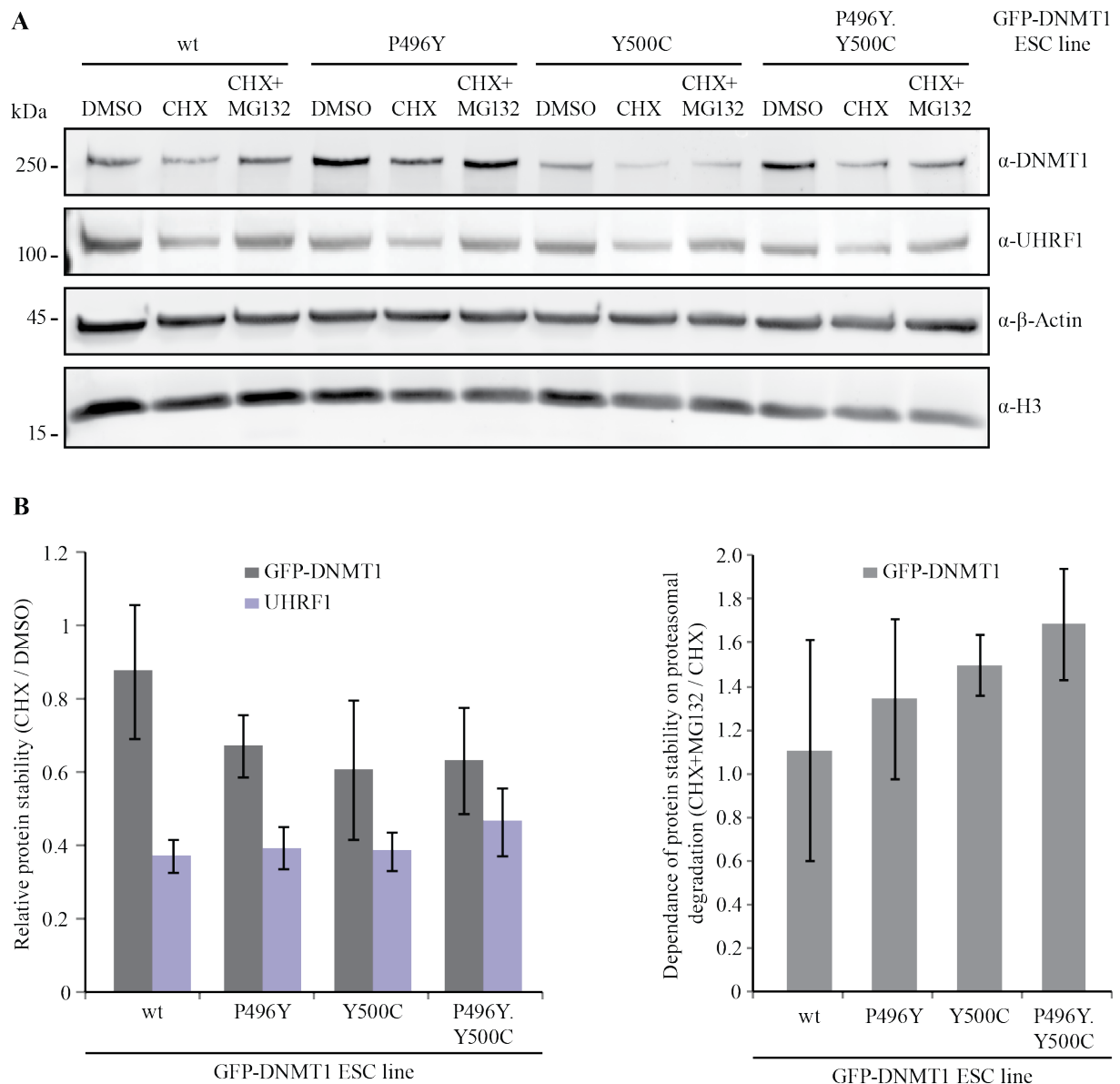


Figure S6. HSANIE associated mutations in the TS domain decrease protein stability of DNMT1. **(A and B)** Cycloheximide (CHX) and MG132 assay in *Dnmt1*^{-/-} ESCs stably expressing GFP-DNMT1 wt or P496Y and Y500C as well as P496Y.Y500C. Cells were treated with 0.03 mg/ml CHX or in addition with 30 μ M MG132 for 5 h. Treatment with DMSO served as mock control. After harvesting, expression levels of GFP-DNMT1 and endogenous UHRF1 were analyzed by immunoblotting with specific antibodies and by quantification using the ImageJ gel analysis tool. β -Actin and H3 staining was used to normalize protein expression levels of DNMT1 and UHRF1. **(A)** Shown is one representative blot of three independent experiments. For immunoblotting, we used specific antibodies for DNMT1, UHRF1, β -Actin and H3. **(B)** Quantification of the relative protein

stability. Ratios of the relative protein expression 5 h after CHX treatment over the relative expression in the mock control (DMSO treatment) were calculated. Shown are mean values \pm SEM of three biological replicates relative to the loading controls (left panel). Dependence of GFP-DNMT1 wt and HSANIE mutants expression on proteasomal degradation (right panel). Ratios of the relative protein expression 5 h after CHX+MG132 treatment over the relative expression after 5 h of CHX treatment were calculated. Ratios > 1 illustrate that protein abundance is higher in the combinatorial treatment (CHX+MG132) when compared to the treatment with CHX suggesting a dependence of protein stability on the proteasomal pathway. Shown are mean values \pm SEM of three biological replicates relative to the loading controls.

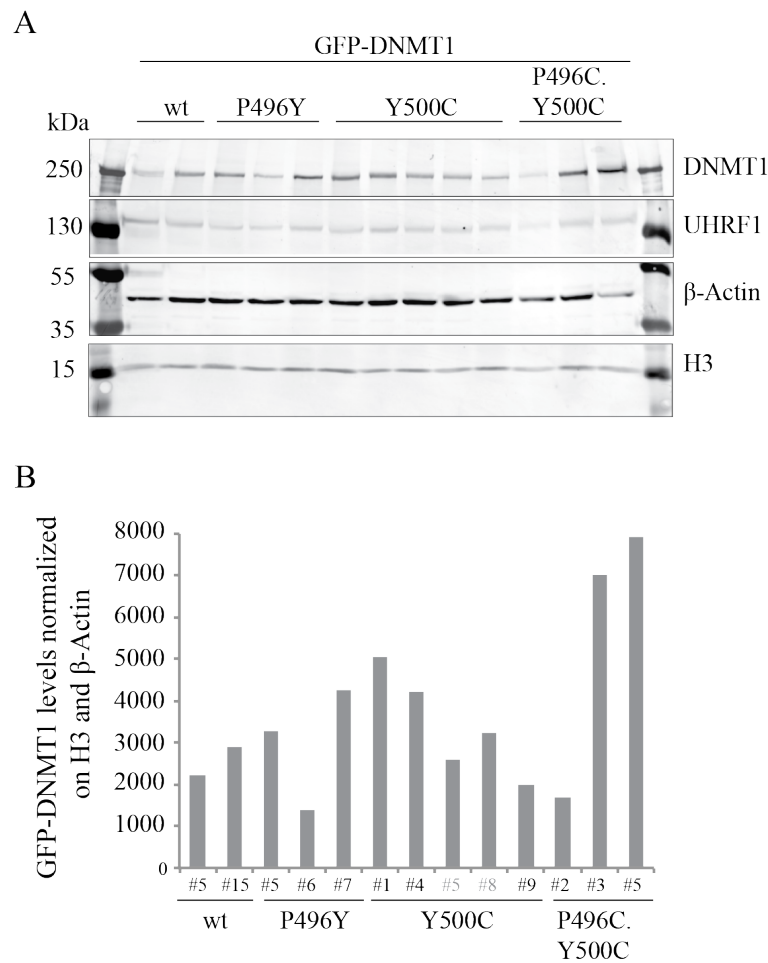


Figure S7. GFP-DNMT1 wt and HSANIE associated point mutations in the TS domain of DNMT1 show different expression levels. Expression levels of GFP-DNMT1 and endogenous UHRF1 were analyzed by immunoblotting with specific antibodies and by quantification of the resulting signals using the ImageJ gel

analysis tool. Equal loading was confirmed by β -Actin and H3 staining and was used to normalize protein expression levels of DNMT1 and UHRF1. **(A)** Shown are the different expression levels of GFP-DNMT1 wt, P496Y, Y500C, and P496Y.Y500C via western blot analysis. For immunoblotting, we used specific antibodies for DNMT1, UHRF1, β -Actin and H3. **(B)** Quantification of protein levels. Signals were quantified with ImageJ. For the neuronal progenitor differentiation, embryonic stem cell clones with different expression levels of GFP-DNMT1 were chosen (clones shown in black).

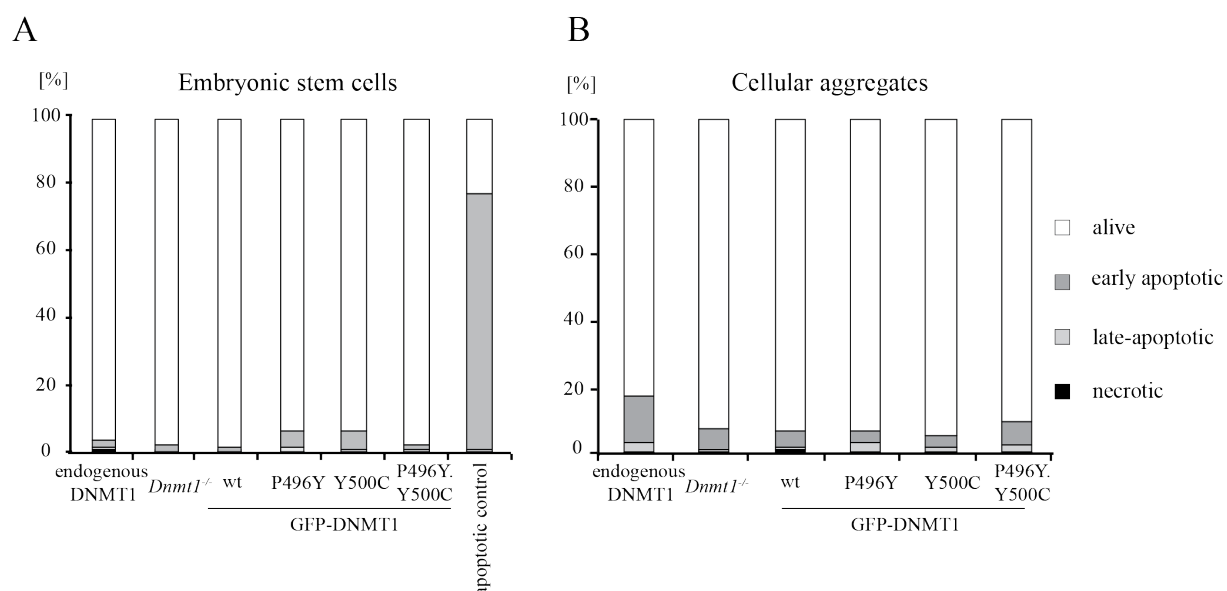


Figure S8. Cell Viability. Flow cytometric analysis of alive, early apoptotic, late apoptotic and necrotic cells in embryonic stem cells **(A)** and dissociated cellular aggregates on day 8 of differentiation **(B)**. Apoptosis was induced by treatment of cells with staurosporine (“apoptotic control”).

Supplementary table

Table S1. List of primers used for real-time quantitative PCR.

Target gene	Primer sequence Forward	Primer sequence Reverse
<i>Gata6</i>	CAAAAGCTTGCTCCGTAACA	GGTCGCTTGTGTAGAAGGAGAAG
<i>Gapdh</i>	CATGGCCTTCCGTGTTCTTA	CTTCACCACCTTCTTGATGTCATC
<i>Brachyury</i>	CTCCAACCTATGCGGACAATTC	ATGACTCACAGGCAGCATGCT
<i>Nestin</i>	ACTCTGCTGGAGGCTGAACT	CAAGGAAATGCAGCTTCAGCTT
<i>Oct4</i>	TCACCCTGGGCGTTCTCTT	GGCCGCAGCTTACACATGTT

Results

2.2 DNA methylation requires a DNMT1 ubiquitin interacting motif (UIM) and histone ubiquitination

DNA methylation requires a DNMT1 ubiquitin interacting motif (UIM) and histone ubiquitination

Weihua Qin^{1,3,4,*}, Patricia Wolf^{1,3,4,*}, Nan Liu^{1,3,4}, Stephanie Link^{1,3,4}, Martha Smets^{1,3,4}, Federica La Mastra^{1,3,5}, Ignasi Forné^{2,3}, Garwin Pichler^{1,3,6}, David Hörl^{1,3,4}, Karin Fellinger^{1,3,7}, Fabio Spada^{1,3,8}, Ian Marc Bonapace⁵, Axel Imhof^{2,3}, Hartmann Harz^{1,3,4}, Heinrich Leonhardt^{1,3,4}

¹Department of Biology II, Ludwig Maximilians University Munich, Großhaderner Str. 2, 82152 Planegg-Martinsried, Germany;

²Adolf-Butenandt Institute, Ludwig Maximilians University Munich, Schillerstr. 44, 80336 Munich, Germany; ³Center for Integrated Protein Science Munich (CIPSM), ⁴Nanosystems Initiative Munich (NIM), ⁵Department of Theoretical and Applied Sciences, University of Insubria, Via Manara 7, 21052 Busto Arsizio (VA), Italy

DNMT1 is recruited by PCNA and UHRF1 to maintain DNA methylation after replication. UHRF1 recognizes hemimethylated DNA substrates via the SRA domain, but also repressive H3K9me3 histone marks with its TTD. With systematic mutagenesis and functional assays, we could show that chromatin binding further involved UHRF1 PHD binding to unmodified H3R2. These complementation assays clearly demonstrated that the ubiquitin ligase activity of the UHRF1 RING domain is required for maintenance DNA methylation. Mass spectrometry of UHRF1-deficient cells revealed H3K18 as a novel ubiquitination target of UHRF1 in mammalian cells. With bioinformatics and mutational analyses, we identified a ubiquitin interacting motif (UIM) in the N-terminal regulatory domain of DNMT1 that binds to ubiquitinated H3 tails and is essential for DNA methylation *in vivo*. H3 ubiquitination and subsequent DNA methylation required UHRF1 PHD binding to H3R2. These results show the manifold regulatory mechanisms controlling DNMT1 activity that require the reading and writing of epigenetic marks by UHRF1 and illustrate the multifaceted interplay between DNA and histone modifications. The identification and functional characterization of the DNMT1 UIM suggests a novel regulatory principle and we speculate that histone H2AK119 ubiquitination might also lead to UIM-dependent recruitment of DNMT1 and DNA methylation beyond classic maintenance.

Keywords: UHRF1; histone ubiquitination; DNMT1; DNA methylation

Cell Research advance online publication 12 June 2015; doi:10.1038/cr.2015.72

Introduction

Epigenetic mechanisms including DNA and histone modifications are crucial for the regulation of gene expression during development. DNA methylation occurs

at the C5 position of cytosine residues, mostly within cytosine-guanine dinucleotides (CpG), and is involved in imprinting, X-chromosome inactivation, stable transcriptional repression, genome stability and tumorigenesis [1]. DNA methylation patterns are established by the *de novo* methyltransferases DNMT3A and DNMT3B during gametogenesis and early development, and are propagated by the maintenance methyltransferase DNMT1 after DNA replication in somatic cells.

DNMT1 comprises a regulatory N-terminal domain (NTD), which covers two-thirds of the molecule, and a C-terminal catalytic domain (CD), which contains all essential motifs of active C5 DNA methyltransferases. The NTD controls the subcellular distribution of DNMT1 during the cell cycle and its enzymatic activity. A sub-domain in the NTD was initially described as a targeting sequence (TS) as it was found to mediate the associa-

*These two authors contributed equally to this work.

Correspondence: Heinrich Leonhardt

E-mail: h.leonhardt@lmu.de

Tel: +49 89 2180 74232; Fax: +49 89 2180 74236

⁶Current address: Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried, Germany

⁷Current address: Intervet International GmbH, Unterschleißheim, Germany

⁸Current address: Department of Chemistry, Ludwig Maximilians University Munich, Germany

Received 21 November 14; Revised 27 March 15; Accepted 7 May 2015

tion of DNMT1 with late replicating pericentromeric heterochromatin [2]. Subsequent studies defined a distinct proliferating cell nuclear antigen (PCNA) binding domain (PBD) responsible for the interaction with the replication machinery [3]. The subnuclear localization of DNMT1 undergoes characteristic changes throughout the cell cycle reflecting PBD-mediated PCNA binding during S phase and TS domain-mediated heterochromatin association during late S and G2 phase [4, 5]. The association of DNMT1 with the replication machinery enhances methylation efficiency, but is not strictly required for postreplicative maintenance DNA methylation [6, 7]. In contrast, the TS domain was found to be required for DNMT1 enzymatic activity [8, 9]. However, the molecular mechanism of TS domain function in the regulation of maintenance DNA methylation remains elusive.

Besides its role in replication-independent heterochromatin binding, the TS domain mediates DNMT1 homodimerization [9] and autoinhibition [10, 11]. A recent crystal structure shows that the TS domain inserts into the DNA binding pocket of the CD, indicating a role of intramolecular interactions in the regulation of DNMT1 activity [10, 11]. Moreover, the TS domain interacts with the SET- and RING-associated (SRA) domain of ubiquitin like, containing PHD and RING finger domains 1 (UHRF1) [12–14]. In contrast to UHRF2, the interaction of UHRF1 with DNMT1 was found to be S phase-dependent [15].

UHRF1, also known as NP95 (mouse) or ICBP90 (human), has been reported as a crucial cofactor for maintenance DNA methylation. Mice lacking UHRF1 show a similar phenotype as *Dnmt1* null (*Dnmt1*^{−/−}) mice that manifests in genomic DNA hypomethylation and developmental arrest at embryonic day 9.5 [16–18]. The SRA domain of UHRF1 preferentially binds to hemimethylated DNA resulting from semiconservative DNA replication and is, therefore, thought to play an important role in loading DNMT1 onto newly synthesized DNA substrates [16, 17, 19–22]. The heterochromatin association of UHRF1 is also mediated by the tandem Tudor domain (TTD), which forms an aromatic cage for specific binding of histone H3 tails containing a trimethylated lysine 9 (H3K9me3) residue [22–25]. The plant homeodomain (PHD) was reported to act in combination with the TTD to read the H3K9me3 mark [26] and to contribute to large-scale reorganization of pericentromeric heterochromatin [27]. In addition, UHRF1 harbors a really interesting new gene (RING) domain endowed with ubiquitin E3 ligase activity *in vitro*, which is required for growth regulation of tumor cells [24, 28]. The ubiquitination state and stability of DNMT1 is controlled by UHRF1 and the ubiquitin-specific protease USP7 [29, 30]. UHRF1 over-

expression leads to DNA hypomethylation by the destabilization and delocalization of DNMT1 [31]. Besides its role in marking DNMT1 for proteasomal degradation, UHRF1 also exerts its ubiquitin E3 ligase activity on histone substrates [24, 25].

A recent study describes replication-dependent H3K23 ubiquitination by UHRF1 in *Xenopus* extracts [32]. Knockdown and rescue experiments in HeLa cells showed that SRA domain-mediated DNA binding as well as RING domain-dependent E3 ubiquitin ligase activity of UHRF1 are required for H3 ubiquitination. Expression of the SRA and RING domain mutants in *Uhrf1*^{−/−} mouse cells could neither restore DNMT1 replication targeting nor DNA methylation levels. A deletion of large parts of the DNMT1 TS domain abolished binding to ubiquitinated H3K23 *in vitro*, but effects on enzymatic activity were not investigated. In particular, the structure and function of the rather large TS domain with its multiple roles and interactions remain to be clarified.

In this study, we elucidate the complex interplay between UHRF1 and DNMT1. While we could confirm the general role of UHRF1 in recruiting DNMT1 to substrate sites by direct interaction, we found that DNMT1 targeting and activities are essentially controlled by specific binding to histone tails ubiquitinated by UHRF1. We generated defined mutations in different UHRF1 domains that retained SRA domain-mediated binding to hemimethylated DNA substrate sites, TTD-mediated recognition of H3K9me3 and binding of DNMT1, but did not allow maintenance DNA methylation. We could show that binding to unmodified H3R2 via the PHD and ubiquitination of H3K18 via the RING domain are required for UHRF1 to mediate maintenance DNA methylation. In turn, we identified a ubiquitin interacting motif (UIM) in the TS domain of DNMT1 that reads this ubiquitin mark and is strictly required for maintenance DNA methylation *in vivo*. These results show the manifold regulatory mechanisms controlling DNMT1 activity and illustrate the multifaceted interplay between DNA and histone modifications.

Results

The interaction of DNMT1 with UHRF1 is required for maintenance DNA methylation

To test whether the interaction of DNMT1 with UHRF1 is indeed required for maintenance DNA methylation, we generated stable cell lines based on *Dnmt1*^{−/−} ESCs expressing green fluorescent protein (GFP) fusions of either DNMT1 wild-type (GFP-DNMT1 wt) or a truncated TS domain deletion mutant (GFP-DNMT1 Δ458–500) that is defective in binding to UHRF1 (Figure

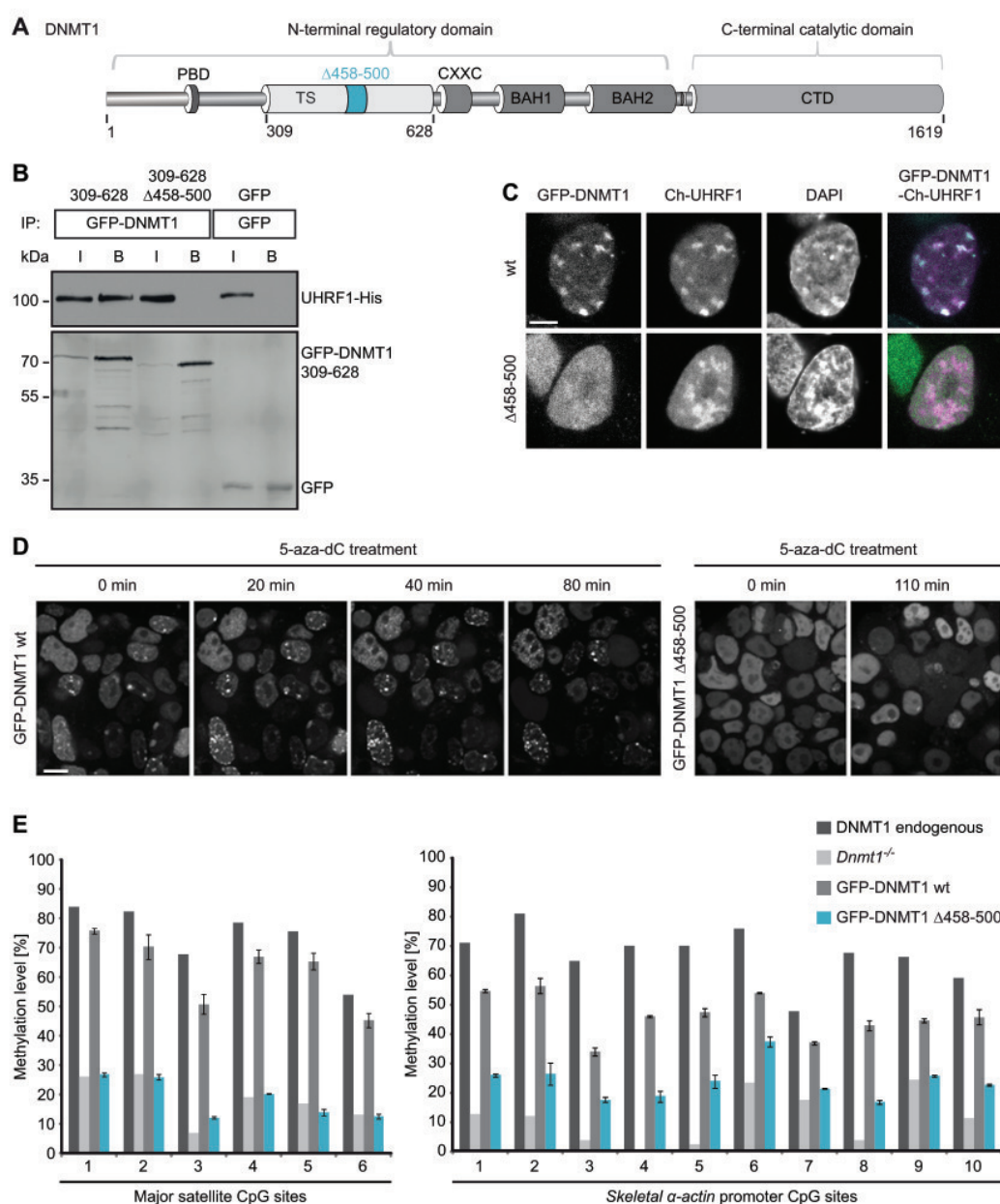


Figure 1 The DNMT1 TS domain is required for UHRF1 interaction, heterochromatin targeting and maintenance DNA methylation. **(A)** Schematic outline of DNMT1 domains and the TS domain deletion (Δ458-500). DNMT1 comprises a large N-terminal domain (NTD) harboring the PCNA binding domain (PBD), the targeting sequence (TS) domain and two bromo adjacent homology (BAH) domains. The active catalytic center of DNMT1 resides within its C-terminal domain (CD). **(B)** Co-immunoprecipitation of UHRF1-His and the GFP-DNMT1 TS domain (309-628) wild-type (wt) or Δ458-500 constructs. Both constructs were co-expressed in HEK 293T cells and after immunoprecipitation of GFP fusions, bound proteins were detected by western blot with an anti-UHRF1 and an anti-GFP antibody. GFP was used as negative control. I, input; B, bound. **(C)** Confocal mid sections of fixed ESCs stably expressing GFP-DNMT1 wt or Δ458-500 mutant constructs. Ch-UHRF1 was transiently co-expressed to illustrate heterochromatic regions, DAPI was used for counterstaining. Scale bar, 5 μm. **(D)** Covalent complex formation of GFP-DNMT1 wt and GFP-DNMT1 Δ458-500 mutant were analyzed by an *in vivo* trapping assay. Confocal mid-sections of ESCs stably expressing GFP-DNMT1 wt and deletion mutant constructs before and after treatment with the mechanism-based inhibitor 5-aza-dC are displayed. Scale bar, 10 μm. **(E)** Local DNA methylation analyses at the major satellite repeats and the *skeletal α-actin* promoter. CpG methylation levels of mouse *Dnmt1*^{-/-} ESCs stably expressing GFP-DNMT1 wt or GFP-DNMT1 Δ458-500 mutant constructs were analyzed by bisulfite treatment of genomic DNA, PCR amplification and direct pyrosequencing. The methylation level of the J1 wt cell line (endogenous DNMT1) and untransfected *Dnmt1*^{-/-} cells are shown for comparison. Mean values ± SD from two different clones were calculated, respectively.

1A and 1B). The deleted region was determined by a sequence alignment of TS domains from higher eukaryotes and a conserved core region of the domain was chosen for mutational analysis (Supplementary information, Figure S1A). In contrast to GFP-DNMT1 wt, GFP-DNMT1 Δ 458-500 did not co-localize with cherry (Ch)-UHRF1 and showed a dispersed distribution in the nucleus (Figure 1C), suggesting that the interaction with UHRF1 is essential for subnuclear localization of DNMT1.

Next, we investigated the role of UHRF1 interaction for the catalytic function of DNMT1. Notably, GFP-DNMT1 Δ 458-500 that did not interact with UHRF1 was able to fully methylate hemimethylated DNA substrates *in vitro* (Supplementary information, Figure S1C). To test the DNA methylation activity of this deletion mutant *in vivo*, we made use of a trapping assay. In this assay, the cytosine analogue 5-aza-2'-deoxycytidine (5-aza-dC) forms an irreversible covalent complex with the methyltransferase at the C6 position of the cytosine residue when incorporated into DNA during replication thereby trapping the enzyme at DNA replication foci. Trapped DNMT1 fractions increase over time and allow monitoring the activity-dependent accumulation of DNMT1 at its target sites [33]. In ESCs stably expressing GFP-DNMT1 wt, foci of immobilized protein emerged already within 20 min (Figure 1D, left panel). In contrast, GFP-DNMT1 Δ 458-500 was not enriched at replication foci even after 110 min, indicating that the deletion mutant is unable to methylate newly replicated DNA in living cells (Figure 1D, right panel). To pursue this idea, we further analyzed site-specific DNA methylation levels of stable GFP-DNMT1 wt and Δ 458-500 ESC lines (Supplementary information, Figure S1B). GFP-DNMT1 could restore local DNA methylation at the major satellite repeats in *Dnmt1*^{-/-} ESCs leading to an average methylation level of 62% that is comparable to the level of the wt cell line expressing the endogenous protein (74%, Figure 1E, left panel). In contrast, the DNMT1 mutant deficient in UHRF1 binding was unable to reestablish local DNA methylation patterns resulting in decreased levels at the major satellite repeats (average 19%) similar to the *Dnmt1*^{-/-} control cell line (average 18%). Consistently, a similar defect of GFP-DNMT1 Δ 458-500 in DNA methylation activity was observed at the single-copy sequence of the *skeletal α -actin* promoter (Figure 1E, right panel). Furthermore, similar results were obtained from DNA methylation analyses at the minor satellite repeats and the *Dnmt1o* promoter confirming that stable expression of GFP-DNMT1 Δ 458-500 could not restore DNA methylation in a *Dnmt1*^{-/-} cell line (Supplementary information, Figure S1D).

In summary, we provide strong evidence that the

GFP-DNMT1 Δ 458-500 mutant deficient in UHRF1 binding, even though able to methylate DNA substrates *in vitro*, cannot restore DNA methylation patterns in *Dnmt1*^{-/-} ESCs. These findings suggest that the interaction of DNMT1 with UHRF1 is required to maintain DNA methylation *in vivo*.

The PHD and RING domain of UHRF1 are essential for maintenance DNA methylation

Cooperative binding of the UHRF1 TTD to di- and trimethylated histone H3K9 and of the SRA domain to hemimethylated DNA was described as a prerequisite for targeting DNMT1 to its substrate and for subsequent DNA methylation [34]. Given the regulatory impact of these two domains, we were interested in how the PHD and RING domain of UHRF1 may functionally contribute to maintenance DNA methylation by DNMT1. To this end, we introduced point mutations in the PHD and RING domain (UHRF1-GFP H346G and UHRF1-GFP H730A, respectively) that are expected to prevent coordination of zinc ions by zinc-finger motifs (Figure 2A and Supplementary information, Figure S2A). Consequently, the mutation in the RING domain significantly reduced the E3 ubiquitin ligase activity of UHRF1 *in vivo* (Supplementary information, Figure S2C and S2D). Notably, the preference of UHRF1-GFP for hemimethylated DNA was not impaired by the PHD and RING domain mutations (Supplementary information, Figure S2B).

First, we tested whether the point mutations in the PHD and RING domain influence the interaction of UHRF1 with DNMT1. UHRF1-GFP wt as well as UHRF1-GFP H346G and UHRF1-GFP H730A still co-precipitated with red fluorescent protein (RFP)-DNMT1, indicating that the mutations do not affect the interaction with DNMT1 directly (Figure 2B). In addition, the unaltered interactions were confirmed by a fluorescent three-hybrid assay [35, 36]. In this assay, UHRF1-GFP fusion constructs were used as baits by tethering them to a *lac* operator (*lacO*) array present in baby hamster kidney (BHK) cells that simultaneously express RFP-DNMT1 as a prey. Accumulation of RFP-DNMT1 at the *lacO* spot enriched for UHRF1-GFP wt, UHRF1-GFP H346G or UHRF1-GFP H730A clearly demonstrates that the mutant proteins were still able to interact with DNMT1 *in vivo* (Figure 2C).

In order to perform functional studies on the PHD and RING domain mutants, we stably expressed GFP-tagged UHRF1 wt, UHRF1 H346G or UHRF1 H730A in *Uhrf1*^{-/-} ESCs. Similar to wt, also UHRF1-GFP H346G and UHRF1-GFP H730A showed focal enrichment at heterochromatin (Figure 2D, first panel and Supplementary information, Figure S2E). Thus, the mutations

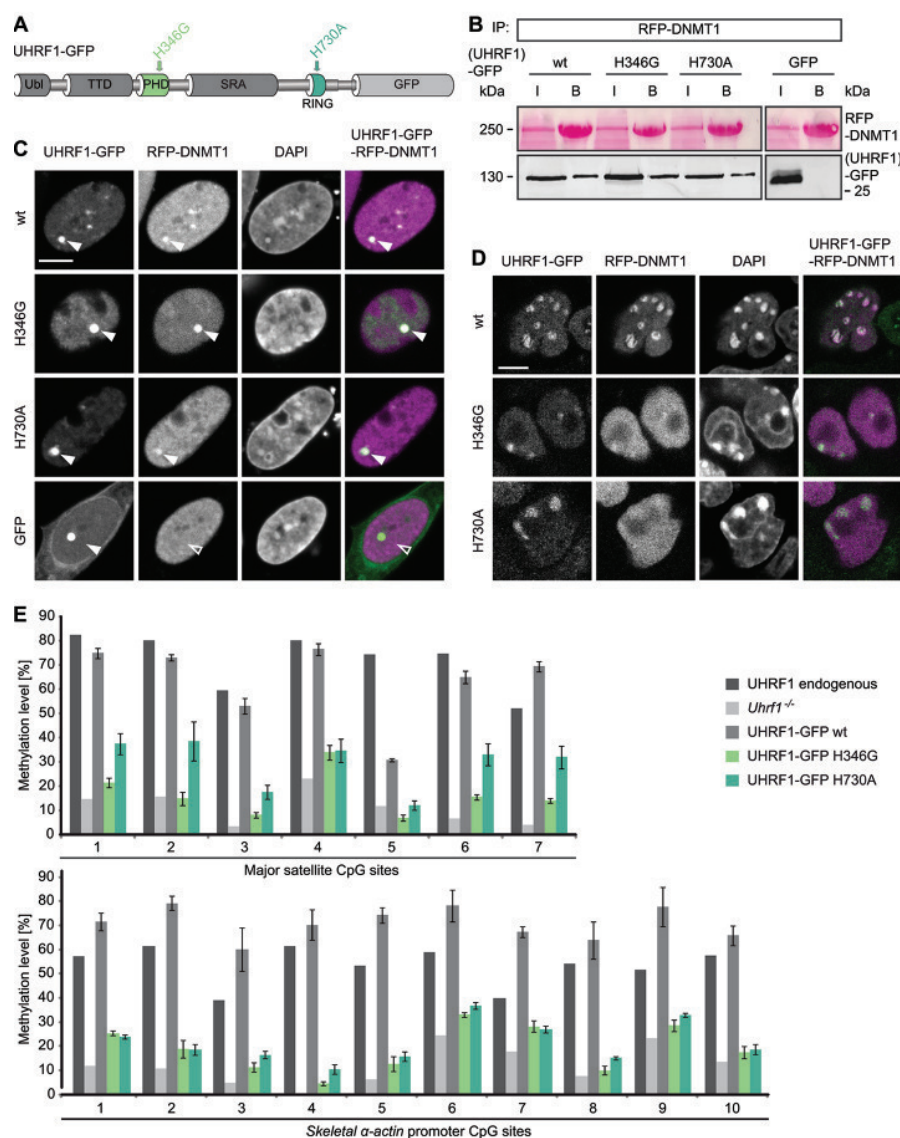


Figure 2 Mutations in the PHD and RING domain of UHRF1 affect DNMT1 targeting and maintenance DNA methylation, but not the interaction with DNMT1. **(A)** Schematic outline of the multidomain protein UHRF1. UHRF1 harbors a ubiquitin-like (Ubl) domain, a plant homeodomain (PHD) and a tandem Tudor domain (TTD) followed by a SET and RING-associated (SRA) domain and a really interesting new gene (RING) domain. UHRF1-GFP expression constructs carrying point mutations in the PHD (H346G) and RING domain (H730A) are illustrated. **(B)** Co-immunoprecipitation of UHRF1-GFP wt or PHD and RING domain mutants co-expressed with RFP-DNMT1 in HEK 293T cells. RFP-DNMT1 was immunoprecipitated using the RFP-Trap and bound UHRF1-GFP was detected by western blot with an anti-GFP antibody. GFP was used as negative control. Immunoprecipitated RFP-DNMT1 is shown by Ponceau staining. I, input; B, bound. **(C)** Fluorescence three-hybrid assay for visualization of the interaction RFP-DNMT1 with UHRF1-GFP wt or PHD and RING domain mutants. Displayed are confocal mid sections of BHK cells carrying a stably integrated Lac-operator array that were triple transfected with LacI fused to the GFP-binder, UHRF1-GFP constructs and RFP-DNMT1. DAPI was used for chromatin counterstaining. Closed arrows indicate the co-localization of both proteins at the *lacO* spot, open arrows indicate no co-localization. GFP was used as negative control. Scale bar, 5 μ m. **(D)** Confocal mid sections of fixed *Uhrf1*^{-/-} ESCs stably expressing UHRF1-GFP wt or PHD and RING domain mutant constructs. RFP-DNMT1 was transiently co-expressed and DNA was counterstained by DAPI. Scale bar, 5 μ m. **(E)** Local DNA methylation analyses at major satellite repeats and the *skeletal α -actin* promoter. CpG site methylation levels of mouse E14 *Uhrf1*^{-/-} ESCs stably expressing UHRF1-GFP wt or PHD and RING domain mutant constructs were analyzed by bisulfite treatment of genomic DNA, PCR amplification and direct pyrosequencing. The methylation level of E14 wt ESCs (endogenous UHRF1) and untransfected E14 *Uhrf1*^{-/-} cells are shown for comparison. Mean values \pm SD from two different clones were calculated, respectively.

do not affect localization of UHRF1. In contrast to its chromatin association in the UHRF1-GFP wt cell line, transiently co-expressed RFP-DNMT1 did not co-localize with UHRF1-GFP H346G and UHRF1-GFP H730A, but showed a dispersed distribution in the nucleus (Figure 2D, second panel). This observation is consistent with the result of a staining for endogenous DNMT1 (Supplementary information, Figure S2F). Only in the UHRF1-GFP wt cell line, endogenous DNMT1 was enriched at S phase-specific replication sites, whereas it was diffusely distributed in the nucleus of the mutant cell lines pointing towards a defective DNMT1 targeting mechanism. To examine if DNMT1 methylation activity depends on the PHD and RING domain of UHRF1, we performed site-specific methylation analyses at heterochromatic regions. Consistent with defects in targeting DNMT1 to replication sites, DNA methylation levels at the major satellite repeats and the *skeletal α -actin* promoter revealed that both UHRF1-GFP H346G and UHRF1-GFP H730A were not able to mediate DNA remethylation by DNMT1 in *Uhrfl*^{-/-} ESCs in contrast to UHRF1-GFP wt (Figure 2E). Especially at the major satellite repeats, the average DNA methylation in the PHD mutant cell lines remained nearly unchanged (16%) from the *Uhrfl*^{-/-} control cell line (11%). Also, the average methylation levels in the RING domain mutant cell lines (29%) did not reach the wt DNA methylation level (62%) at the major satellite repeats. Similar results were obtained for the minor satellite repeats and the *Dnmt1* promoter (Supplementary information, Figure S3A). Consistent with this site-specific DNA hypomethylation, the stable UHRF1 mutant cell lines also showed decreased global DNA methylation levels as compared with the wt (Supplementary information, Figure S3B and S3C). Partial rescue of global DNA methylation in the RING domain mutant cell lines could be due to residual E3 ubiquitin ligase activity of UHRF1-GFP H730A (Figure 3B, Supplementary information, Figure S2C and S2D).

To exclude the possibility that DNA hypomethylation might result from lower expression of the PHD and RING domain mutant (Figure 3A), we performed a transient rescue assay in *Uhrfl*^{-/-} ESCs. Even though expression levels of the mutant constructs exceeded those of the UHRF1-GFP wt, the PHD and RING domain mutants could not mediate remethylation at the major satellite repeats (Supplementary information, Figure S3D and S3E) arguing for functional rather than expression defects.

In summary, the PHD and RING domain mutants, although not affecting UHRF1 heterochromatin localization or the direct interaction with DNMT1, cannot mediate either targeting of DNMT1 to replication foci nor maintenance DNA methylation. These findings suggest

that these UHRF1 domains contribute to the recruitment of DNMT1 by indirect mechanisms.

The PHD and RING domain of UHRF1 are required for ubiquitination of histone H3

Histone H3 has been reported as a UHRF1-dependent ubiquitination target in *Xenopus* egg extracts [32], providing a potential mechanism for the recruitment of DNMT1 to chromatin. Thus, we set out to investigate whether H3 ubiquitination required PHD-mediated histone binding and RING domain-mediated ubiquitin E3 ligase activity of UHRF1 in mammalian cells. To this end, we extracted histones from wt or *Uhrfl*^{-/-} ESCs and detected modified H3. As expected, histone H3 was less ubiquitinated in the absence of UHRF1 (Figure 3A and 3B), indicating that UHRF1 serves as a ubiquitin E3 ligase for H3 in mammalian cells. We also found that ubiquitination levels of histone H3 in *Uhrfl*^{-/-} ESCs stably expressing the RING domain mutant UHRF1-GFP H730A were not rescued to the level of wt cells. Surprisingly, the PHD mutant UHRF1-GFP H346G also could not restore ubiquitination of histone H3 (Figure 3A and 3B).

Since the PHD has been reported to bind to unmodified H3R2 [26, 37–39], we investigated the role of this histone residue in H3 ubiquitination by mutational analyses. Compared with GFP-H3 wt, ubiquitination of a GFP-H3 R2A mutant expressed in human embryonic kidney (HEK) 293T cells was clearly reduced (Figure 3C) pointing towards an important role of the R2 residue for UHRF1-dependent H3 ubiquitination.

To further test the histone binding properties of the PHD mutant *in vitro*, we performed a peptide pull-down assay with wt or PHD and RING domain mutant UHRF1-GFP using H3 peptides with an unmodified, trimethylated or acetylated K9 residue. The mutation in the RING domain did not alter the histone binding of UHRF1-GFP showing a preference for unmodified and K9 trimethylated H3 peptides similar to the wt protein (Figure 3D). The mutation in the PHD, however, decreased the binding to both, the unmodified and the K9 trimethylated peptide. We further examined the histone binding preferences of UHRF1-GFP with an *in vitro* histone tail binding assay. The results revealed the binding of UHRF1-GFP to unmodified but not R2 dimethylated H3 histone tails (Supplementary information, Figure S4), consistent with prior K_d measurements [39]. As the PHD of UHRF1 has been shown to bind unmethylated H3R2 residues and to contribute to the K9 methylated H3 histone binding of the TTD [26, 37–39], we propose that PHD-dependent histone binding is required for UHRF1-mediated ubiquitination of histone H3.

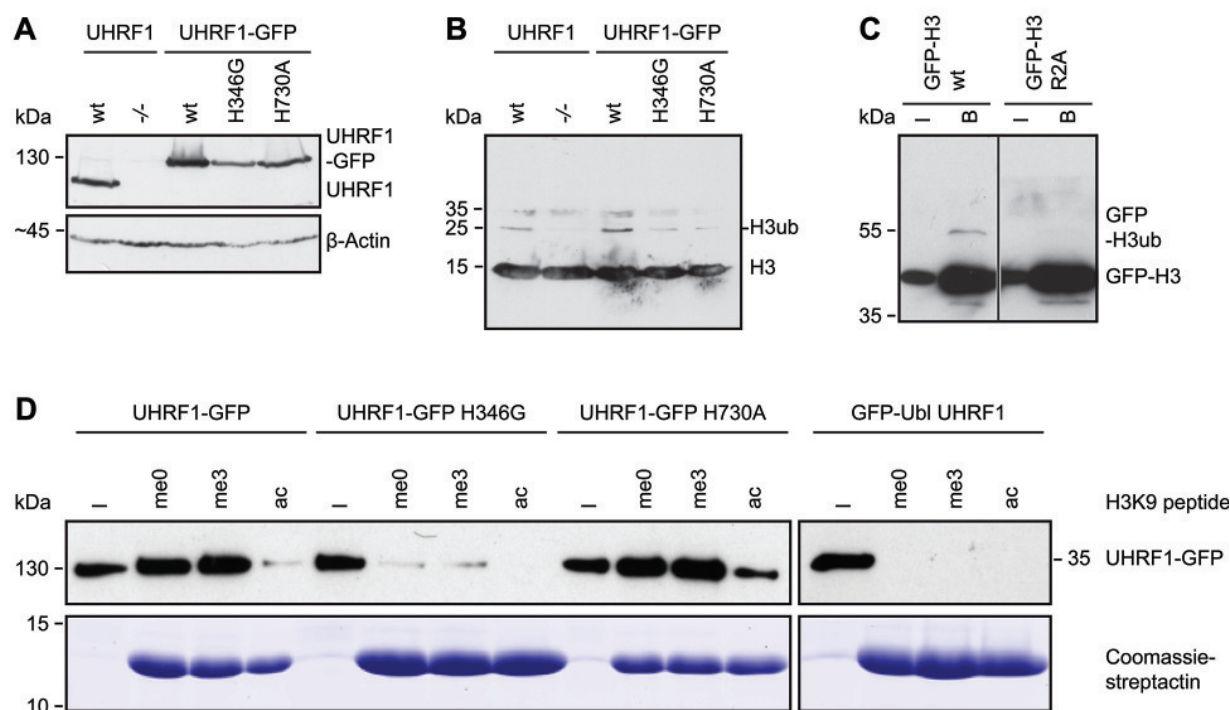


Figure 3 Histone H3 ubiquitination requires the UHRF1 PHD and RING domain. **(A)** Western blot analyses of endogenous UHRF1 or stably expressed UHRF1-GFP wt or H346G and H730A mutants in E14 *Uhrf1*^{-/-} ESCs with an anti-UHRF1 antibody. Equal loading is shown by an anti- β -Actin antibody. **(B)** Analyses of H3 ubiquitination from acid extracted histones derived from the different cell lines in (A). A specific anti-H3 antibody was used for detection. **(C)** Ubiquitination of GFP-H3 in dependence on R2. GFP-H3 wt and the arginine to alanine mutant (R2A) were co-expressed with UHRF1-His in HEK 293T cells, respectively, and after immunoprecipitation with the GFP-Trap, the bound fraction was detected by western blotting with a specific anti-H3 antibody. I, input; B, bound. **(D)** *In vitro* peptide pull-down assay of UHRF1-GFP wt or the PHD and RING domain mutants from crude cells extracts of HEK 293T cells using H3 peptides (amino acid 1-20) that were either unmodified (me0), K9 trimethylated (me3) or K9 acetylated (ac) and functionalized on streptactin beads. The GFP-Ubl domain of UHRF1 was used as negative control, Coomassie-stained streptactin is shown as loading control. I, input.

UHRF1 ubiquitinates histone H3 on K18 in mammalian cells

Using *Xenopus* extracts immunodepleted for DNMT1, H3 was shown to be ubiquitinated at the K23 residue [32]. To map ubiquitination sites on histone H3 tails in mammalian cells, we performed mass spectrometry using human and mouse cells. In contrast to the results from *Xenopus* extracts, the K18 residue of histone H3 was identified as novel ubiquitination site in mouse ESCs, while the K23 residue was unmodified or acetylated (Figure 4A and 4B). Relative quantification of H3 peptides containing ubiquitinated K18 and an unmodified or acetylated K23 residue showed a reduction of K18 ubiquitination in ESCs lacking UHRF1 (Figure 4C and 4D). Similarly, immunoprecipitation of GFP-UHRF1 from HEK 293T cells and subsequent mass spectrometry also revealed ubiquitination at K18 but not at K23 (Supplementary information, Figure S5A). Comparison of ubiquitination levels of overexpressed GFP-H3 carrying R2A, K18A or

K23A mutations suggests that in this constellation K23 could also be modified (Supplementary information, Figure S5B). Interestingly, the GFP-H3 R2A construct showed reduced ubiquitination levels indicating that the R2 residue plays a role in regulating H3 ubiquitination.

DNMT1 harbors a UIM that mediates binding to ubiquitinated H3 and is essential for DNA methylation activity in vivo

To unravel how H3 ubiquitination may contribute to maintenance DNA methylation, we screened DNMT1 for potential binding motifs. With bioinformatics analyses, we identified a ubiquitin interacting motif (UIM) in the N-terminal regulatory domain of DNMT1. This motif is located in a region spanning from amino acid 380 to 399 of mouse DNMT1 and shows striking similarity to UIMs of known ubiquitin interacting proteins (Figure 5A). Comparison of the ubiquitin binding properties between GFP-DNMT1 wt and mutants either lacking

the UIM ($\Delta 356$ -404) or containing substitutions of the relevant and conserved amino acids in the motif to alanine (D381A-E382A-S392A, D381A-E382A-M385A-S392A-D395A, Figure 5A and Supplementary information, Figure S6A) showed a defect in the association with ubiquitinated histone H3 and ubiquitinated H2AK119 (Figure 5B, 5C and Supplementary information, Figure S6B-S6D). To further elucidate UIM-dependent ubiquitinated histone binding, we quantified modified H318-26 peptides bound by GFP-DNMT1 wt or the UIM mutants by mass spectrometry. Whereas H3 histone peptides ubiquitinated at K18 and acetylated or unmodified at K23 co-immunoprecipitated with GFP-DNMT1 wt, only little to no ubiquitinated peptide signals were detected for the UIM mutants (Figure 6A, 6B). GFP-DNMT1 $\Delta 458$ -500 defective in UHRF1 interaction (Figure 1B) showed reduced (Figure 6B) or undetectable (Figure 5B, 5C) binding to ubiquitinated H3 and H2A. This deletion located in a TS domain region C-terminal of the UIM might affect the integrity and functionality of the motif responsible for ubiquitin binding. Therefore, we cannot rule out that apart from disrupted UHRF1 binding also defects in the association with ubiquitinated histones contributed to the observed changes in subnuclear distribution and protein function of GFP-DNMT1 $\Delta 458$ -500 (Figure 1C-1E).

Besides a decreased binding to ubiquitinated H3, the TS domain point and deletion mutants exhibited an increased binding to H3 or core histones compared with GFP-DNMT1 wt (Figure 5B). Therefore, specific binding of DNMT1 to ubiquitinated H3 via its UIM might prevent the enzyme from stable chromatin association and thereby facilitate DNA methylation.

To clarify the functional role of the UIM in maintenance DNA methylation *in vivo*, we performed a functional complementation assay in *Dnmt1*^{-/-} ESC lines transiently expressing GFP-DNMT1 wt, GFP-DNMT1 $\Delta 356$ -404, GFP-DNMT1 D381A-E382A-S392A or GFP-DNMT1 D381A-E382A-M385A-S392A-D395A. Local DNA methylation analyses at the major satellite repeats and the *skeletal α -actin* promoter showed that the UIM mutants were not able to reestablish DNA methylation patterns (Figure 6C). GFP-DNMT1 wt restored DNA methylation at the major satellite repeats to 48%. By comparison, the UIM deletion and point mutants were not able to rescue resulting in low average methylation levels of 20% to 23% comparable to untransfected *Dnmt1*^{-/-} ESCs (15%). Similar results were also observed at the minor satellite repeats and the *Dnmt1*o promoter (Supplementary information, Figure S7A).

Given that the GFP-DNMT1 TS UIM deletion and point mutants were able to interact with Ch-UHRF1 (Supplementary information, Figure S7B), we were in-

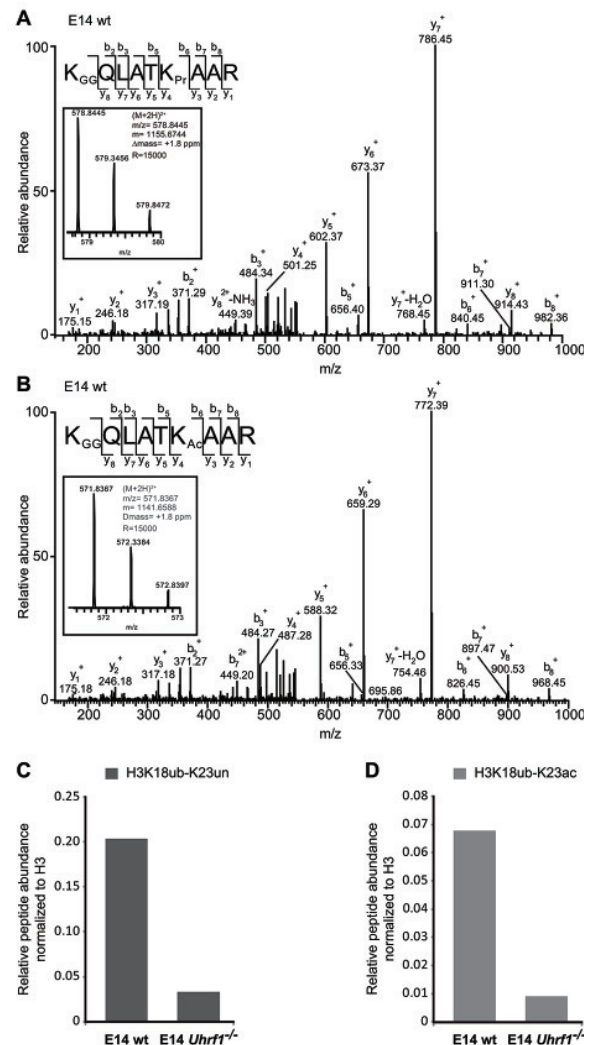


Figure 4 UHRF1 ubiquitinates histone H3 at K18 in mammalian cells. **(A)** Identification of H3 18-26 peptides carrying ubiquitination (GG) at K18 and no modification (Pr) at K23 by LC-MS/MS. MS2 fragmentation spectrum of the precursor ion is shown in the inset. An almost complete series of b and full y product ions generated by CID fragmentation were detectable providing a high confidence in its correct identification and localization of the ubiquitin modification. Inset: mass, charge and measurement error determination of the H3 18-26 peptides K18GGK23Pr in the E14 wt sample. Displayed is the isotopic distribution of the H3 peptide from which the mass to charge ratio (*m/z*), the charge (2+) and the monoisotopic mass value (*m*) were derived. Δm : difference between the expected and the measured masses; *R*: resolution of the MS measurement. **(B)** Identification of H3 18-26 peptides carrying ubiquitination (GG) at K18 and acetylation (Ac) at K23 by LC-MS/MS as in (A). **(C, D)** Quantification of H3 18-26 peptides carrying ubiquitination (ub) at K18 and an unmodified (un) or acetylated (ac) K23 residue from E14 wt and E14 *Uhrf1*^{-/-} samples. Extracted ion chromatograms of the ions corresponding to the peptides of interest were used for the quantification. The signals were normalized against the total amount of analyzed H3 proteins.

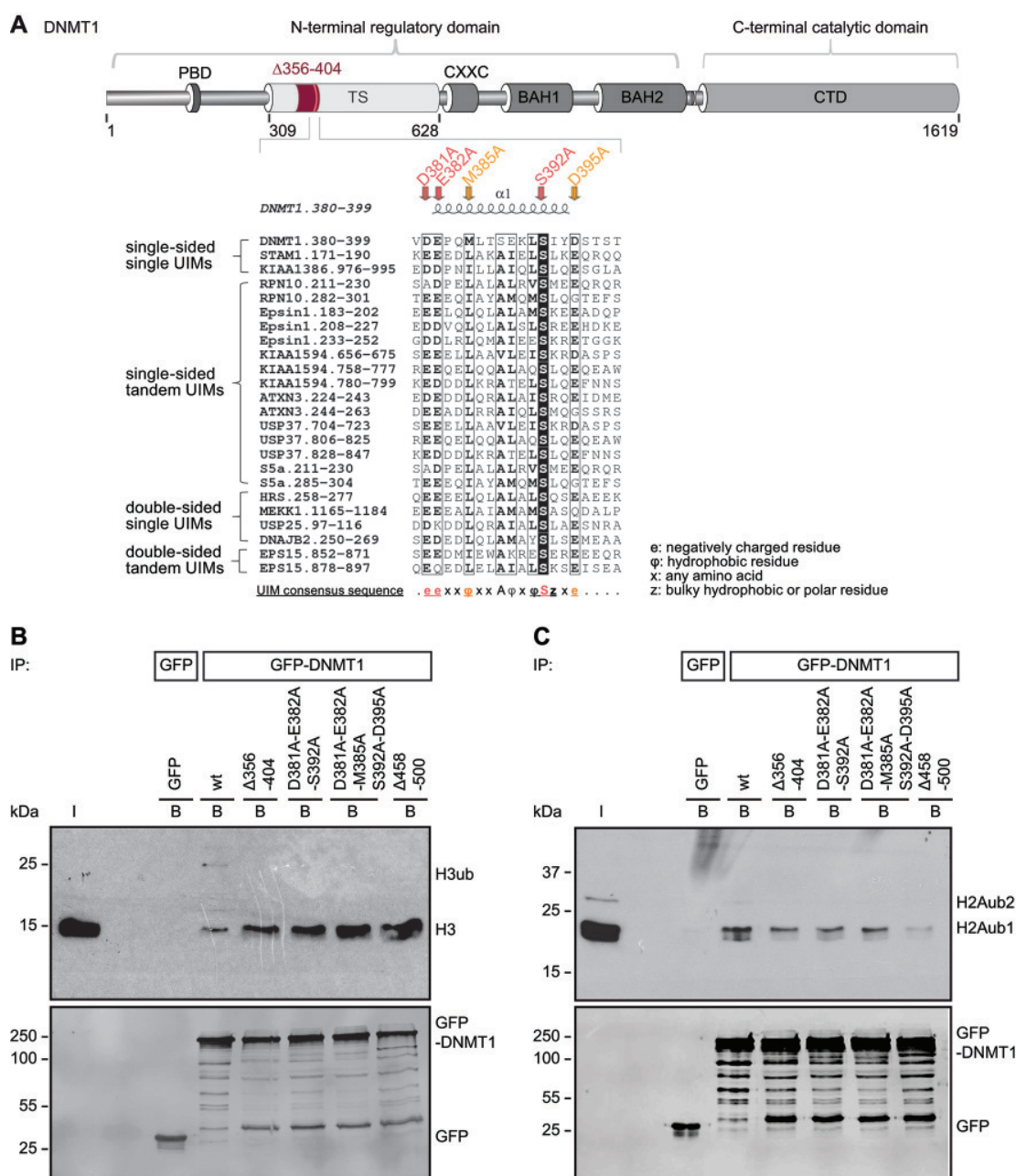


Figure 5 The TS domain of DNMT1 harbors a ubiquitin interacting motif (UIM) that is essential for binding to ubiquitinated H3 and H2A. **(A)** Schematic outline of the UIM in the TS domain of DNMT1 and indication of the UIM deletion ($\Delta 356-404$) and the point mutations (D381A-E382A-S392A and D381A-E382A-M385A-S392A-D395A). A peptide sequence of DNMT1 encompassing amino acid 380-399 was aligned with peptide sequences of proteins previously known to contain UIMs. Identical amino acids are highlighted in black, highly similar amino acids are framed in black. The secondary structure of the DNMT1 region (pdb: 3EPZ [10]) harboring the UIM is displayed on top of the sequence alignment generated using ESPript [78]. The consensus sequence for single-sided UIMs [58] is shown below. The UIMs were found by scanning the protein primary sequences against a collection of motifs in ExPASy Prosite. Putative subgroups of UIMs are indicated on the left. **(B)** Ubiquitinated histone H3 binding assay. After extraction of histones from HEK 293T cells, the extracts were incubated with GFP-DNMT1 wt or mutants immobilized on the GFP-Trap and the bound fractions were analyzed by western blotting with specific anti-H3 and anti-GFP antibodies. GFP was used as negative control. I, input; B, bound. **(C)** Ubiquitinated histone H2A binding assay as in (B). Bound fractions were analyzed by western blotting with specific anti-H2AK119ub and anti-GFP antibodies. Analyses of the anti-H2AK119ub antibody specificity and of peptides isolated from the corresponding band are shown in Supplementary information, Figure S6C and S6D. H2Aub1, monoubiquitinated H2A; H2Aub2, diubiquitinated H2A.

terested in how the UIM in DNMT1 has an influence on the subnuclear localization of the protein. Immunostaining of replicating DNA with a specific anti-PCNA antibody indicated that GFP-DNMT1 wt was enriched at S phase-specific replication foci, while GFP-DNMT1 Δ 356-404, GFP-DNMT1 D381A-E382A-S392A and GFP-DNMT1 D381A-E382A-M385A-S392A-D395A showed only weak association with the PCNA-stained replication sites especially in late S phase (Supplementary information, Figure S8). To analyze the UIM-dependent enrichment of DNMT1 at late-replicating heterochromatin, we quantified mean fluorescence intensities at chromocenters compared with the nucleoplasmic region (Figure 7A). In late S phase ES and mouse embryonic fibroblasts (MEF) cells, GFP-DNMT1 wt localized at chromocenters, whereas the UIM mutations abolished heterochromatin enrichment (Figure 7B and 7C). These results clearly demonstrate the key role of the UIM in DNMT1 targeting via ubiquitinated histone H3 binding and for maintenance DNA methylation in mammalian cells.

Discussion

DNA methylation is an important epigenetic modification regulating gene expression in development and disease. A key question is how methylation marks are set, maintained and removed. According to previous models, DNA methylation marks are set by the *de novo* methyltransferases DNMT3A and DNMT3B during development and maintained by the maintenance DNA methyltransferase DNMT1 that specifically recognizes and modifies hemimethylated DNA substrates. However, the preference of DNMT1 for hemimethylated DNA measured *in vitro* [40–43] is not sufficient to explain efficient maintenance of DNA methylation patterns over many cell division cycles *in vivo*. The interaction of DNMT1 with the replication protein PCNA was shown to enhance maintenance DNA methylation by a factor of two, but not to be essential [6, 7]. In contrast, the interacting factor UHRF1 recruiting and allosterically activating DNMT1 is essential for DNA methylation [14, 16, 17, 44]. In this study, we have now dissected the distinct role of different UHRF1 and DNMT1 domains in directing DNA methylation.

In line with previous studies, we show that, albeit being weak, the TS domain-mediated interaction of DNMT1 with the SRA domain of UHRF1 is required for targeting and function of DNMT1 *in vivo*. Accordingly, truncated DNMT1 (Δ 458–500) deficient in UHRF1 binding showed weaker association with chromocenters in late S phase mouse fibroblasts [4] and failed to maintain DNA methylation in ESC (Figure 1).

Heterochromatin binding of UHRF1 is mediated by

the TTD, PHD and SRA domain and defects in any of these three domains lead to decreased DNA methylation by DNMT1 [34, 45, 46]. Accordingly, it was postulated that UHRF1 reads and binds repressive histone marks and hemimethylated DNA and via direct protein-protein interaction recruits DNMT1 for maintenance DNA methylation.

Defects of a RING domain mutant (C713A, C515A and C716A) in restoring ubiquitinated H3 in HeLa cells after knockdown of human DNMT1 and UHRF1 have previously been reported [32]. We found that the RING domain, though not directly involved in UHRF1 chromatin binding or interaction with DNMT1, is indispensable for DNA methylation by DNMT1. Remarkably, a UHRF1 RING domain mutant (H730A) with diminished ubiquitin E3 ligase activity (Supplementary information, Figure S2C and S2D) that could still bind DNMT1 (Figure 2B), hemimethylated DNA and K9 trimethylated H3 peptides *in vitro* (Figure 3D and Supplementary information, Figure S2B) and chromocenters *in vivo* (Supplementary information, Figure S2E), nonetheless failed in recruiting DNMT1 to replication sites (Figure 2D and Supplementary information, Figure S2F). These findings suggest that DNMT1 recruitment to replication forks is not based on direct interaction with UHRF1, but on the catalytic activity of the RING domain. Previously, the RING domain of UHRF1 has been reported to have an autoubiquitination activity [28] and, in addition, to ubiquitinate DNMT1 [29, 30] and histone substrates [24, 25]. A recent study describes that ubiquitination of H3 by UHRF1 provides docking sites for DNMT1 on chromatin and thus couples maintenance DNA methylation and replication [32]. While we could confirm the essential role of UHRF1, we obtained new insights into the complex functional interplay of UHRF1 and DNMT1 domains.

First, in contrast to ubiquitination at K23 in *Xenopus* egg extracts [32], our mass spectrometry results identified H3K18 as ubiquitination target of UHRF1 in mammalian cells (Figure 4A, 4B and Supplementary information, Figure S5A). By mutational analysis in HEK 293T cells, we found that in absence of K18, the mutated GFP-tagged H3 might be ubiquitinated at K23 (Supplementary information, Figure S5B). However, by semiquantitative analysis of endogenous ubiquitinated H3 peptides in wt versus *Uhrf1*^{−/−} mouse ESCs using mass spectrometry, we clearly show the specificity of K18 ubiquitination by UHRF1 and its reduction by UHRF1 depletion (Figure 4C, 4D). Second, in the previous study, a deletion of 100 amino acids within the DNMT1 TS domain (Δ 325–425) caused a loss of histone binding *in vitro* [32]. The TS domain is, however, involved in multiple interactions and required for proper folding, stability and activity of DNMT1. The incomplete structural information indicates

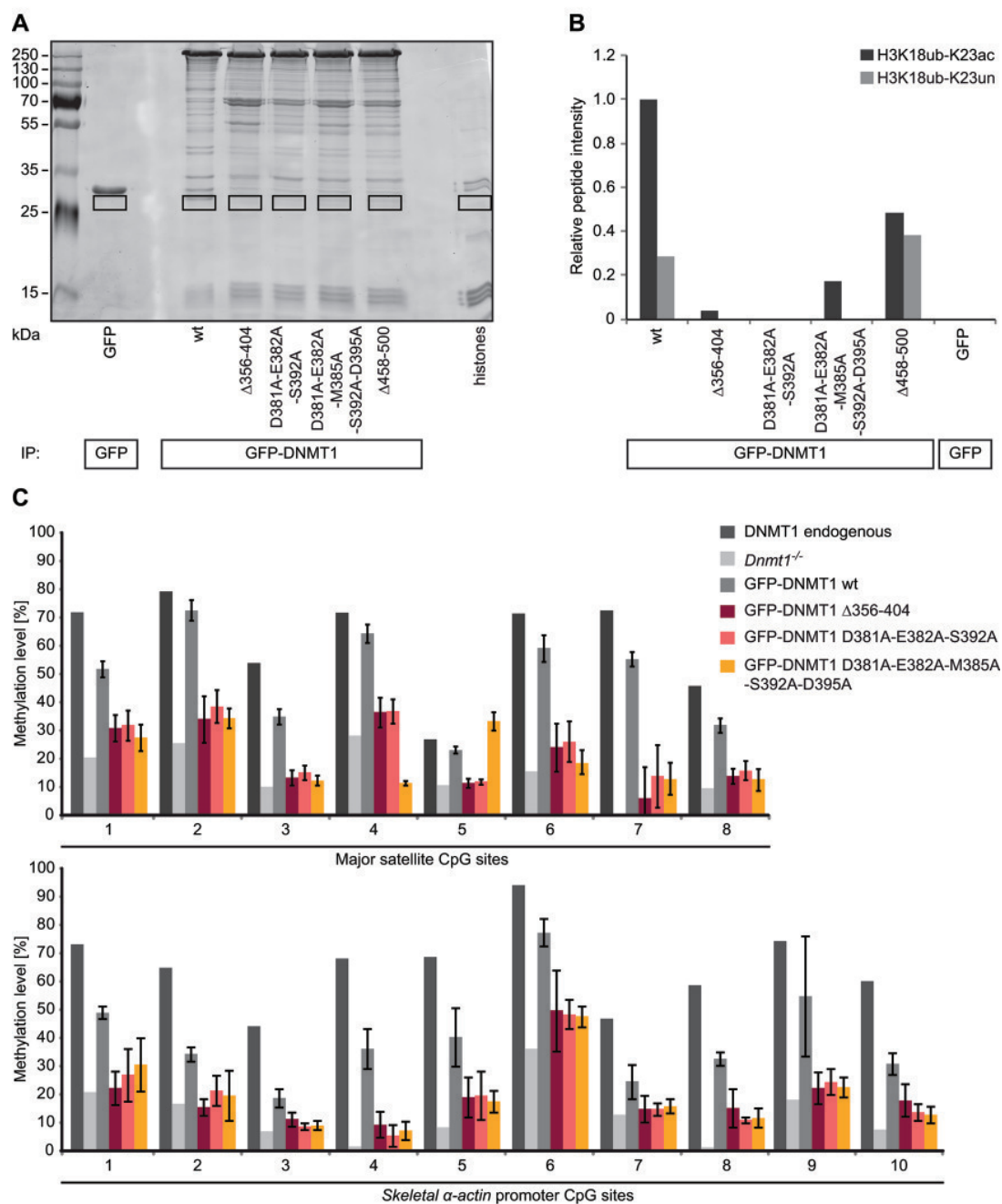


Figure 6 The DNMT1 UIM is required for ubiquitinated H3K18 binding and for DNA methylation. **(A)** Ubiquitinated histone binding experiments using GFP-DNMT1 wt or UIM mutants as well as the $\Delta 458-500$ mutant deficient in binding to UHRF1. Equal amounts of GFP fusions were immobilized on the GFP-Trap and incubated with acid extracted histones. Bound proteins were visualized by Coomassie staining and the fractions highlighted by black rectangles were analyzed by mass spectrometry. GFP was used as negative control. **(B)** Quantification of H3 18-26 peptides carrying ubiquitination (ub) at K18 and an acetylated (ac) or unmodified (un) K23 residue from histone binding experiment shown in (A). Extracted ion chromatograms of the ions corresponding to the peptides of interest were used for quantification (H3K18ubK23ac: $m/z = 571.8353 \pm 10$ ppm; H3K18ubK23un: $m/z = 578.8441 \pm 10$ ppm). **(C)** Local DNA methylation analyses of J1 *Dnmt1*^{-/-} ESCs expressing GFP-DNMT1 wt or $\Delta 356-404$ and UIM point mutants. CpG methylation levels at the major satellite repeats and the *skeletal α -actin* promoter were analyzed by bisulfite treatment of genomic DNA, PCR amplification and direct pyrosequencing. Methylation levels of untransfected J1 *Dnmt1*^{-/-} cells are shown for comparison. Mean values \pm SD from three to four biological replicates were calculated, respectively.

different TS domain conformations and a role in auto-inhibition of the CD, but does not provide any further mechanistic insights [10, 11, 47]. With bioinformatics and mutational analyses, we identified a conserved UIM located in the TS domain of DNMT1 (amino acids 381–395) that mediates the recognition of ubiquitinated H3 *in vitro* (Figure 5, 6A, 6B and Supplementary information, Figure S6B). Localization and activity analyses with specific mutants *in vivo* clearly indicated that the UIM is required for DNMT1 subnuclear distribution and maintenance DNA methylation (Figures 6C, 7 and Supplementary information, Figure S7A and S8).

Last, we could show that besides hemimethylated DNA binding by the SRA domain [32], UHRF1 PHD binding to H3R2 is also required for H3 ubiquitination and subsequent DNA methylation (Figure 2E and 3B). Therefore, we propose that cooperative chromatin binding of the TTD, the PHD and the SRA domain constitutes a prerequisite for H3K18 ubiquitination. These ubiquitinated histone tails are recognized by the UIM and thus mediate DNMT1 chromatin binding. Thereby, UHRF1 acts as a reader and writer of histone marks and via recruitment of DNMT1 dynamically links DNA and histone modification pathways. Based on these results, we propose a ubiquitination-dependent chromatin targeting mechanism for DNMT1 that is essential for maintenance DNA methylation after replication (Figure 8A). The identification and functional characterization of a UIM in DNMT1 not only changes our view of maintenance DNA methylation, but also opens new perspectives for the involvement of DNMT1 in other repressive epigenetic pathways (Figure 8B).

Besides association with ubiquitinated H3, we found that DNMT1 also binds ubiquitinated H2AK119 (Figure 5C and Supplementary information, Figure S6C, S6D). Consistently, DNMT1 was recently detected among proteins binding to H2A ubiquitinated at K118 in *Drosophila*, corresponding to K119 in mammals [48]. H2AK119 ubiquitination is catalyzed by RING1A/1B, two components of the Polycomb repressive complex 1 (PRC1), and plays an important role in regulating gene expression [49]. Similar to UHRF1-dependent H3 ubiquitination, H2A ubiquitination by RING1A/1B might also contribute to DNA methylation. We speculate that UIM-mediated binding of DNMT1 to ubiquitinated H2AK119 might direct DNMT1 to un- or hemimethylated sites dependent on PRC1 ubiquitination activity (Figure 8B, left half).

PRC1-dependent H2A ubiquitination further leads to PRC2 recruitment and subsequent H3K27 methylation [50]. Enhancer of Zeste homolog 2 (EZH2), a component of PRC2, writes methylated H3K27 and interacts with DNMTs. This interaction was shown to be required for DNA methylation of EZH2 target promoters [51].

DNMT1 depletion in differentiated cells affects H2A ubiquitination-dependent PRC2 recruitment at pericentromeric heterochromatin [52]. Thus, UIM binding to ubiquitinated H2A is likely DNA replication independent and DNMT1 might function as adaptor protein mediating PRC2 recruitment and repressive Polycomb domain formation.

Besides recruiting DNMT1 to specific sites on chromatin, the UIM could also play a role in the allosteric activation of the enzyme. The UIM is located within the TS domain of DNMT1 that had been shown to bind the CD and thereby inhibit catalytic activity [10, 11]. It is tempting to speculate that competitive UIM binding to ubiquitinated histone tails displaces the TS domain from the DNA binding pocket and abolishes autoinhibition of DNMT1.

Given the emerging role of ubiquitination in DNA methylation, it is interesting to notice that ubiquitination is a highly dynamic post-translational modification that can be reversed by ubiquitin-specific proteases (USPs). The UHRF1-DNMT1 complex has been reported to contain USP7 that deubiquitinates and stabilizes DNMT1 [29, 30]. Thus, USP7 might in addition modulate the ubiquitination status of histone H3 and thereby regulate DNMT1 association with chromatin. An alternative pathway controlling DNMT1 chromatin association could involve the recently described chromatin acetylation of H3K18 and K23 [53, 54]. Acetylated H3K18 is enriched at the transcriptional start sites of active and poised genes [55]. Thus, H3K18 acetylation might counteract ubiquitination and thereby prevent binding and silencing of active genes by DNMT1. The dynamic interplay of ubiquitination and acetylation of H3K18 likely controls DNMT1 chromatin binding and thereby directs methylation activity. Studies of UHRF1 and DNMT1 complex composition in different cell cycle phases and cell types should provide further insights into the fine-tuning of DNMT1 activity *in vivo*.

Given the complex role of the large TS domain on the one hand and the scarce structural and mechanistic data on the other hand, our identification of a well defined UIM provides a concrete basis for functional insights. Ubiquitin binding proteins with defined UIMs have been described in various cellular processes like, e.g., sorting of ubiquitinated membrane proteins for lysosomal degradation. The crystal structure of the signal transducing adaptor molecule 1 (STAM1) [56] suggests that three central amino acids in the UIM, L176, A179 and S183 form a hydrophobic interface for ubiquitin binding [57]. Similar to the UIM in STAM1, the UIM in DNMT1 also harbors a conserved hydrophobic amino acid M385 and S392 flanked by negatively charged amino acids (D381, E382 and D395), which we found to be essential in our analyses (Figures 5, 6 and 7). Different from other UIMs,

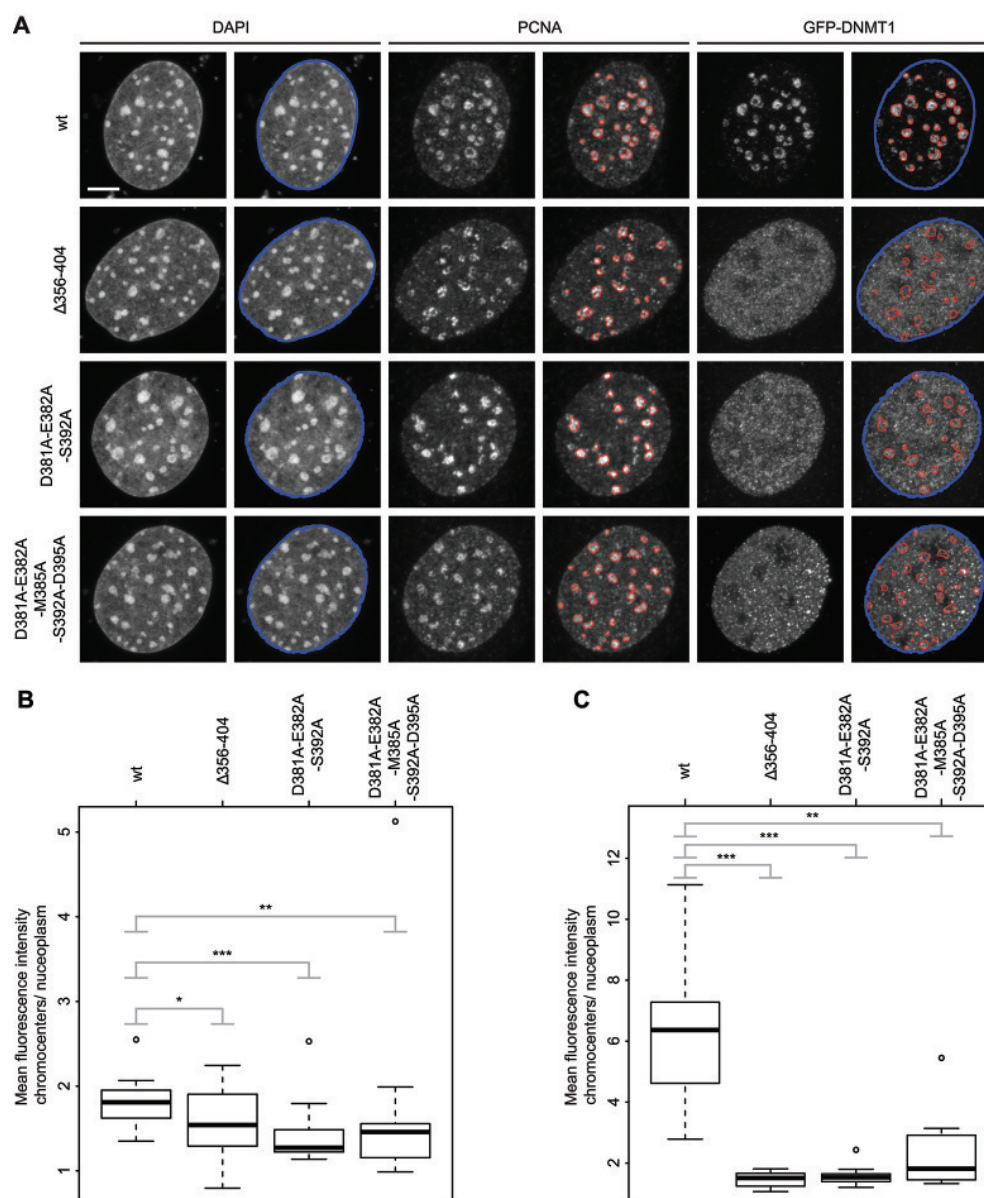


Figure 7 GFP-DNMT1 UIM mutants show a decreased association with PCNA-stained replication sites in late S phase compared with the wt. **(A)** Maximum intensity projections of MEF cells transiently expressing GFP-DNMT1 wt or UIM mutants. Replicating DNA was stained with a specific anti-PCNA antibody and chromatin was counterstained with DAPI. Replication foci masks (red) match the enrichment of GFP-DNMT1 wt in late S phase, whereas the UIM mutants do not show a focal enrichment. Segmentations were generated in an automated fashion using a machine learning algorithm (WEKA). The nuclear mask outlined in blue was based on the DAPI staining, whereas the replication foci masks outlined in red were based on the PCNA staining. Both masks were superimposed on the GFP channels. The GFP-DNMT1 signal inside the red masks (chromocenters) relative to the remainder of the nucleus (nucleoplasm) was quantified. Scale bar, 5 μm . **(B)** Quantification of chromocenter association of GFP-DNMT1 wt or UIM mutants in late S phase J1 *Dnmt1*^{-/-} ESCs. The ratio of the mean GFP fluorescence intensity at chromocenters over the mean intensity in the nucleoplasm is shown in the box plot from 15 (wt), 16 ($\Delta 356-404$), 12 (D381A-E382A-S392A) or 18 (D381A-E382A-M385A-S392A-D395A) cells. The results were further analyzed in R using a Wilcoxon test and considered as statistical significant for $P < 0.05$ (*) and $P < 0.01$ (**) or highly significant for $P < 0.001$ (***). The following P values were calculated: $\Delta 356-404$: $P = 0.049$, D381A-E382A-S392A: $P = 0.0016$ and D381A-E382A-M385A-S392A-D395A: $P = 0.0056$. **(C)** Quantification of chromocenter association of GFP-DNMT1 wt or UIM mutants in late S phase MEF cells as in (B). Eleven (wt), 12 ($\Delta 356-404$, D381A-E382A-S392A) or 10 (D381A-E382A-M385A-S392A-D395A) cells were analyzed. The following P values were calculated in R using a Wilcoxon test: $\Delta 356-404$: $P = 0.00000148$, D381A-E382A-S392A: $P = 0.00000148$ and D381A-E382A-M385A-S392A-D395A: $P = 0.0012$.

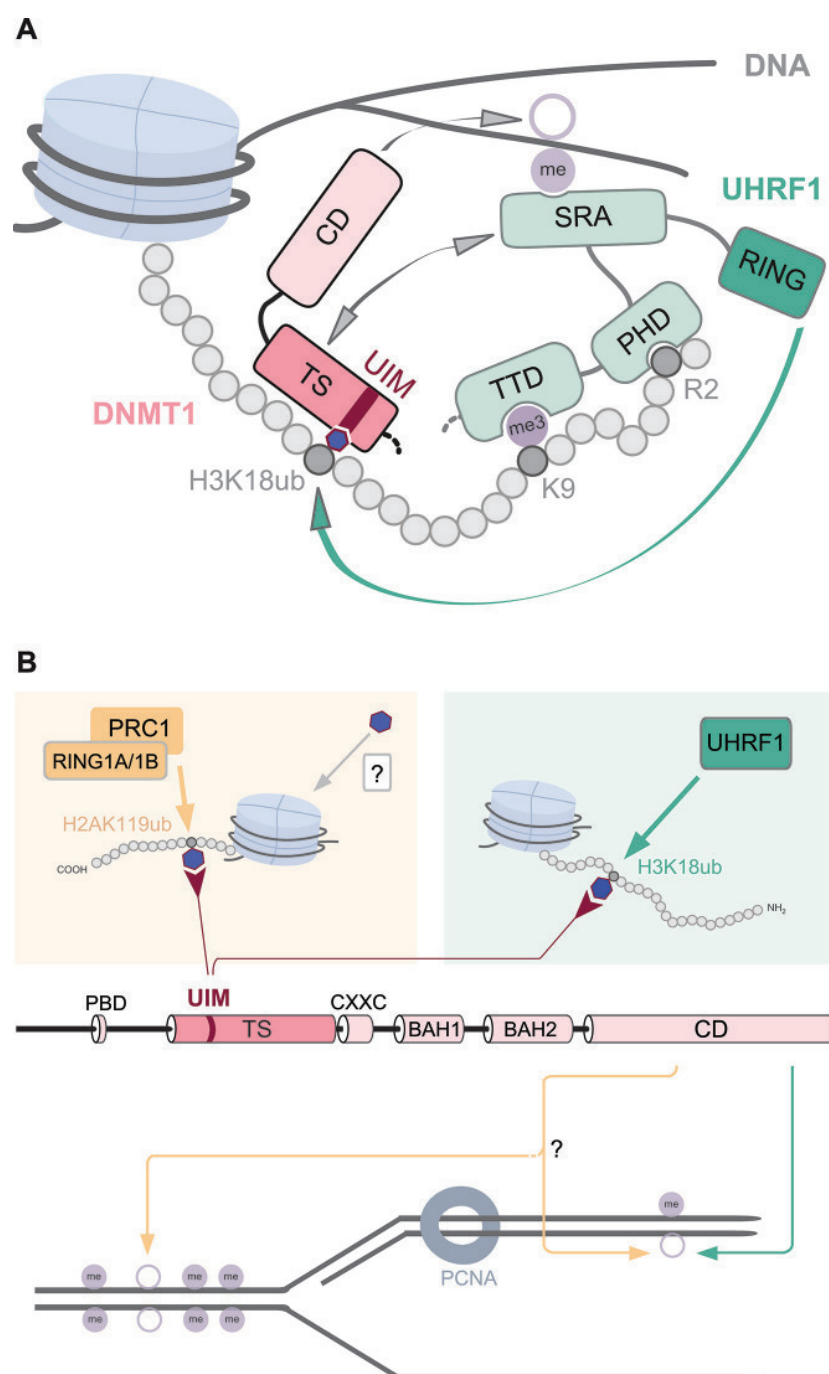


Figure 8 Overview of interactions and modifications controlling DNMT1 activity. **(A)** UHRF1 is enriched at H3 tails as a result of the PHD-mediated binding to H3R2, the TTD-mediated binding to methylated H3K9 and recognition of hemimethylated CpG sites via the SRA domain. By interaction of the SRA domain with the TS domain, DNMT1 is directly recruited to its target sites. UHRF1 chromatin binding via its TTD, PHD and SRA domain is a prerequisite for subsequent H3 ubiquitination by the RING domain. The UIM of DNMT1 binds to H3 tails ubiquitinated at K18 by UHRF1 and is essential for DNMT1 targeting and DNA methylation *in vivo*. **(B)** The previously described direct interaction of DNMT1 with UHRF1 and PCNA is not sufficient for maintenance DNA methylation. Besides the UHRF1-dependent H3K18 ubiquitination recruiting DNMT1 via its UIM for maintenance DNA methylation (right half), we propose an alternative pathway that involves H2AK119 ubiquitination by RING1A/1B of PRC1 (left part). The identification of the DNMT1 UIM now opens the possibility that ubiquitination of histone tail residues by ubiquitin E3 ligases might constitute alternative pathways for DNA methylation by DNMT1 CD beyond classic maintenance. Blue hexagons represent a ubiquitin moiety.

the central conserved A residue is not present in DNMT1 (Supplementary information, Figure S6A). Based on sequence alignments and structural information, UIMs can be subdivided in single-sided single UIMs, as in STAM1, and in single-sided tandem UIMs, as in the proteasome subunit S5a [58] (Figure 5A). The tandem UIMs in S5a provide a model for the recognition of polyubiquitin chains [59]. In contrast, a double-sided single UIM in the hepatocyte growth factor-regulated tyrosine kinase substrate (HRS) allows for efficient binding of multiple monoubiquitinated receptors in the process of endosomal protein sorting [58]. Comparison with these known UIMs suggests that the motif in DNMT1 belongs to the group of single-sided single UIMs, which would be compatible with the recognition of single ubiquitinated histone tails.

In summary, the functional analysis of UHRF1 domains and the identification of a UIM in DNMT1 challenge traditional views of maintenance DNA methylation as a simple copying mechanism. Instead, DNA methylation by DNMT1 requires reading of H3R2, H3K9me3 and hemimethylated DNA by UHRF1 and subsequent ubiquitination of H3K18 by its RING domain thereby integrating signals from different epigenetic pathways. These multiple layers controlling DNMT1 activity suggest that overall methylation densities in chromatin domains are maintained rather than specific methylation patterns precisely copied. The functional characterization of the UIM further raises the possibility that other ubiquitin E3 ligases like RING1A/1B of PRC1 might direct DNMT1 activity to repressive chromatin domains beyond simple maintenance.

Materials and Methods

Expression constructs and antibodies

Fusion constructs were generated using enhanced GFP, monomeric RFP or monomeric Ch. The expression constructs for GFP, RFP-DNMT1, GFP-DNMT1 wt, GFP-DNMT1 Δ 458-500, GFP-DNMT1 309-628 (GFP-TS) and UHRF1-His have been described previously [9, 29, 33, 60, 61]. GFP-TS Δ 458-500, GFP-DNMT1 Δ 356-404 and GFP-DNMT1 point mutant (D381A-E382A-S392A and D381A-E382A-M385A-S392A-D395A) expression constructs as well as UHRF1-GFP H346G and H730A were derived from the corresponding wt constructs by overlap extension PCR [62]. The GFP-UHRF1 single-domain construct for the ubiquitin-like domain (Ubl) was generated by PCR using the corresponding wt full-length construct. Ch-UHRF1 and GFP-UHRF1 expression constructs have been described previously [22, 63]. Expression constructs for GFP-H3 R2A, K18A, K23A as well as K18A-K23A were obtained by overlap extension PCR on the corresponding wt construct. The construct for LacI-GBP has been reported before [36, 64, 65]. All constructs were verified by DNA sequencing (MWG Biotech).

For immunofluorescence staining of heterochromatin, a mouse anti-H3K9me3 and an anti-H4K20me3 antibody were used (Active

Motif). Endogenous DNMT1 was stained with the rat monoclonal antibody 5A10 [4] and PCNA with the rat monoclonal antibody 16D10 [66]. As secondary antibodies an anti-mouse Alexa Fluor 594 and anti-rat Alexa Fluor 647 antibody were applied, respectively (Invitrogen).

For detection of GFP fusion proteins by western blot, a mouse anti-GFP (Roche) or a rat anti-GFP (Chromotek) antibody was used. RFP or Ch fusion proteins were detected by the rat anti-red antibody 5F8 [67]. UHRF1 was visualized by a rabbit anti-UHRF1 antibody [24] and HA-ubiquitin by the mouse monoclonal anti-HA antibody 12CA5. Equal loading of cell lysates was assessed by a mouse anti- β -Actin antibody (Sigma-Aldrich). The rabbit anti-H3 antibody was purchased from Abcam and the anti-H2AK119ub from New England Biolabs. Depending on the expected intensity of the signals, secondary antibodies either conjugated to horseradish peroxidase (anti-rabbit (Biorad), anti-rat and anti-mouse (Dianova)) or conjugated to fluorescent dyes (anti-mouse and anti-rat Alexa Fluor 647 as well as anti-rat Alexa Fluor 488 (Invitrogen)) were applied. For detection of HRP-conjugated antibodies, an ECL Plus reagent (GE Healthcare, Thermo Scientific) was used.

Cell culture, transfection and immunofluorescence staining

HEK 293T and BHK cells were cultured in DMEM supplemented with 10% fetal calf serum and 50 μ g/ml gentamycin (PAA). MEF cells were cultured in DMEM supplemented with 15% fetal calf serum, 0.1 mM β -mercaptoethanol (Invitrogen), 2 mM l-glutamine, 1 \times MEM non-essential amino acids, 100 U/ml penicillin and 100 g/ml streptomycin (PAA). ESCs including J1 wt, *Dnmt1*^{-/-}, E14 wt and *Uhrfl*^{-/-} were cultured without feeder cells in gelatinized flasks as described [33]. Culture medium was supplemented with 1 000 U/ml recombinant leukemia inhibitory factor (Millipore). The *Dnmt1*^{-/-} ESCs used in this study are homozygous for the c allele [68]. Mouse E14 wt and *Uhrfl*^{-/-} cells have been reported before [61]. Mouse ESCs and MEF cells were transfected with FuGENE HD (Roche), Lipofectamine® 2 000 or 3 000 reagent (Invitrogen) according to the manufacturer's instructions. HEK 293T cells and BHK cells were transfected using polyethylenimine as transfection reagent (Sigma) according to the manufacturer's instructions. Cell fixation and microscopy were carried out as described [35].

Generation of stable ESC lines and DNA methylation analyses

Forty-eight hours after expression of GFP-tagged constructs in *Dnmt1*^{-/-} or *Uhrfl*^{-/-} ESCs, GFP-positive mouse ESCs were separated using a fluorescence-activated cell sorting (FACS) Aria II instrument (Becton Dickinson). Stably expressing cells were expanded in selection medium containing 10 μ g/ml blasticidin (GFP-DNMT1 wt and GFP-DNMT1 Δ 458-500) or 500 ng/ml puromycin (UHRF1-GFP wt, H346G and H730A) and GFP-positive cells were FACS sorted a second time. Furthermore, the UHRF1-GFP wt, H346G and H730A cell lines were single-cell sorted. Single clones of GFP-DNMT1 Δ 458-500 and corresponding wt [29] were picked manually. For all cell lines, clones with low expression levels were chosen for further analyses. The level and the accuracy of the expressed GFP fusion constructs were checked by western blot analyses (Figure 3A and Supplementary information, Figures S1B and S3D). For functional analyses of GFP-DNMT1 wt and GFP-DNMT1 UIM mutants (Δ 356-404, D381A-E382A-S392A and D381A-E382A-M385A-S392A-D395A) as well as UHRF1-

GFP wt or UHRF1-GFP point mutants (H346G and H730A) by transient rescue assays, 48 h after expression of these proteins in *Dnmt1*^{-/-} or *Uhrf1*^{-/-} ESCs, respectively, GFP-positive cells were collected with FACS. Genomic DNA isolation, bisulfite conversion and PCR conditions were described before [6, 60, 69]. Primer sets used for amplification of minor satellites, major satellites, *skeletal α -actin* and the *Dnmt1* promoter are listed in Supplementary information, Table S1. All PCR products were analyzed by pyrosequencing (Varionostic), which results in a quantitative data set for individual CpG sites [70].

Co-immunoprecipitation and western blotting

For co-immunoprecipitation assays, the GFP and RFP, Ch or His fusion constructs were co-expressed in HEK 293T cells and protein extracts were normalized to the same GFP or RFP concentration prior to co-immunoprecipitation with the GFP-Trap or RFP-Trap (Chromotek). Bound fractions were first detected by fluorescence intensity measurements and second by western blot analyses.

Acid extraction and TCA precipitation of histones

Histones were isolated by acid extraction as reported previously [71]. In brief, 10⁷ mouse ESCs or HEK 293T cells were treated in hypotonic buffer (10 mM Tris-HCl pH 8, 10 mM KCl, 1.5 mM MgCl₂, 1 mM DTT and 1× Protease Inhibitor, 2 mM PMSF) for 30 min and centrifuged at 1 000× *g* at 4 °C to get the intact nuclei. After washing steps, nuclei were resuspended in 0.4 N H₂SO₄ and incubated on a rotator at 4 °C overnight. After centrifugation, histones in the supernatant were transferred into a fresh reaction tube and precipitated using 33% trichloroacetic acid (TCA). After washing twice with cold acetone, histones were dissolved in H₂O. Histone concentrations were measured using the PierceTM 660 nm protein assay kit (Thermo Scientific).

Ubiquitinated histone binding experiment

For ubiquitinated histone binding experiment, acid extracted histones from HEK 293T cells were used. GFP-DNMT1 and its mutants were immobilized on the GFP-Trap (Chromotek) and incubated with equal amounts of acid extracted histones for 30 min at 4 °C. After washing steps, the bound fractions were analyzed by western blot.

Due to unspecific binding of histones to the eppendorf tubes, we used eppendorf tubes with low binding affinity during mass spectrometry sample preparation.

Immunoprecipitation of ubiquitinated GFP-H3

GFP-H3 wt and R2A mutant constructs were co-expressed in HEK 293T cells with UHRF1-His and harvested after treatment with 2 mM N-ethylmaleimide (NEM, AppliChem) for 5 min. Lysates were prepared by firstly isolating nuclei in hypotonic buffer (10 mM HEPES pH 7.9, 10 mM KCl, 0.1 mM MgCl₂, 10% glycerol, 0.1 mM EDTA, 0.1 mM DTT, 1× protease inhibitor, 2 mM PMSF, 0.1% NP-40, 0.625 mg/ml NEM) and secondly by lysis of the nuclei in hypertonic buffer (20 mM HEPES pH 7.9, 150 mM KCl, 1.5 mM MgCl₂, 10% glycerol, 0.1 mM EDTA, 1 mM DTT, 1× protease inhibitor, 2 mM PMSF, 1 mg/ml DnaseI (AppliChem), 0.625 mg/ml NEM). Prior to immunoprecipitation, the GFP concentration was equalized using lysates from UHRF1-His transfected HEK 293T cells for dilution. After immunoprecipitation of GFP-H3 with the GFP-Trap (Chromotek) and washing (20 mM

HEPES pH 7.9, 300 mM KCl, 10% glycerol, 0.1% Triton X-100), the bound fraction was analyzed by western blot.

For semiquantitative analysis of the GFP-H3 wt or K18A, K23A, K18A-K23A and R2A ubiquitination, the GFP fusion constructs were co-expressed with HA-ubiquitin in HEK 293T cells and 2 days after transfection, the cells were harvested as described above and further processed as reported previously [29].

F3H assay and trapping assay

The F3H assay was performed as described previously [65]. In the trapping assay, mouse ESCs stably expressing GFP-DNMT1 wt or Δ 458-500 were cultured in Ibidi chambers and incubated with 10 μ M of the cytosine analogue 5-aza-2'-deoxycytidine (Sigma). Images were acquired with a UltraVIEW VoX spinning disc microscope (PerkinElmer) assembled to an Axio Observer D1 inverted stand (Zeiss) and using a 63×/1.4 NA Plan-Apochromat oil immersion objective.

In vitro peptide pull-down assay

The peptide pull-down assay from nuclear cell extracts of HEK 293T cells expressing UHRF1-GFP fusion constructs was performed as described [72] with the following modifications. C-terminally biotinylated histone peptides were purchased from PSL and are listed in Supplementary information, Table S2. Streptactin beads (Iba) were used for the immobilization of biotinylated peptides in binding buffer (10 mM Tris-HCl, pH 7.5, 300 mM NaCl, 0.5 mM EDTA, 1 mM DTT). After the binding reaction, beads were washed four times with wash buffer (20 mM HEPES pH 7.9, 20% glycerol, 0.2 mM EDTA, 300 mM KCl, 0.1% Triton X-100). Bound fractions were eluted by boiling in 2× Laemmli sample buffer and analyzed by western blot.

Mass spectrometry

The gel was stained with Coomassie and H3 bands were manually excised, propionylated and digested with trypsin as described before [73] with minor modifications. For peptide extraction, gel slices were incubated twice with 50 μ l of 50% acetonitrile 0.25% TFA and twice more with 50 μ l of acetonitrile. The resulting liquid containing the digested peptides was totally evaporated, redissolved with 15 μ l of 0.1% formic acid and stored at -20 °C until further processing.

Tryptic peptides were injected (5 μ l) in an Ultimate 3 000 HPLC system (LC Packings Dionex). Samples were desalted online in a C18 microcolumn (300 m i.d. × 5 mm, packed with C18 PepMapTM, 5 μ m, 100 Å by LC Packings), and peptides were separated with a gradient from 5% to 60% acetonitrile in 0.1% formic acid over 40 min at 300 nl/min on a C18 analytical column (75 μ m i.d. × 15 cm, packed with C18 PepMapTM, 3 μ m, 100 Å by LC Packings).

The effluent from the HPLC was directly electrosprayed into a linear trap quadrupole-Orbitrap XL mass spectrometer (Thermo Fisher Scientific). The MS instrument was operated in data-dependent mode. Survey full-scan MS spectra (from *m/z* 300–2 000) were acquired in the Orbitrap with resolution *R* = 60 000 at *m/z* 400 (after accumulation to a “target value” of 500 000 in the linear ion trap). The six most intense peptide ions with charge states between two and four were sequentially isolated to a target value of 10 000 and fragmented by collision-induced dissociation and recorded in the linear ion trap. For all measurements with the Orbitrap detector, three lock-mass ions were used for internal calibration [74].

Typical MS conditions were spray voltage, 1.5 kV; no sheath and auxiliary gas flow; heated capillary temperature, 200 °C; normalized collision-induced dissociation energy 35%; activation $q = 0.25$; and activation time = 30 ms.

Mascot 2.3.02 was used for protein identification with the following settings: Database: Swissprot 57.7; Taxonomy: *Homo sapiens* (human); MS tolerance: 10 ppm; MS/MS tolerance: 0.5 Da; peptide FDR: 0.1; protein FDR: 0.01; minimum peptide length: 5; and variable modifications: propionyl (K, N-term), GlyGly (K).

Quantification of modified H3 18-26 peptides was based on the intensities of the MS1 peaks. The spectra depicted in Figure 4A and 4B were used to determine the exact masses ($m/z \pm 10$ ppm) and used as a reference for further quantification.

Quantitative analysis of DNMT1 subnuclear localization

During late S phase, DNMT1 is enriched in replication foci at chromocenters. In order to quantify the subnuclear distribution of GFP-DNMT1 wt and defined UIM mutants the following procedure was used: confocal z-stacks (0.21 μm interval) were acquired with identical scan settings in three color channels to visualize replication foci (anti-PCNA staining, 594 nm excitation), DNMT1 localization (GFP-DNMT1 fusions enhanced with GFP-booster (Chromotek), 488 nm excitation) and DNA counterstaining (DAPI, 405 nm excitation). For each color channel, maximum intensity projections were calculated and only GFP-expressing cells were analyzed. Segmentation of replication foci or whole nuclei was performed with the Weka segmentation plugin [75] in Fiji [76]. Training of the classifier was finalized until the result matched the visual impression (Figure 7A). Due to variations in ESC samples, replication foci were segmented using different classifiers for wt or the different UIM mutants. In contrast, for all somatic cells, one classifier was sufficient to segment replication foci. Whole nuclei were segmented by a classifier based on the DAPI signal. After Weka segmentation, the resulting binary masks were filtered using the particle analyzer of Fiji with a circularity value ≥ 0.25 . To select for cells in late S phase, only replication foci ≥ 150 pixel were further analyzed in the ESC samples. In MEF cell samples, only late S phase cells were imaged and analyzed without applying size exclusion for replication foci. Nuclear masks (size $\geq 3\,000$ pixel) were used to quantify the total amount of GFP fusion protein in a single nucleus. Nuclei were further subsegmented by replication foci masks. For each nucleus, the ratio between the mean GFP signals in replication foci relative to the mean GFP signal outside the foci was calculated. Raw data were corrected for background signals by subtracting the modal grey value. Ratios from all nuclei expressing GFP-DNMT1 wt or UIM mutants were visualized as box plots. Numerical calculations and statistical analysis were performed with R [77].

Statistical analysis

Results were expressed as mean values \pm SD or as mean values \pm SEM from the number of biological replicates indicated in the corresponding figure legend.

Acknowledgments

We thank Stefan Jentsch (Max Planck Institute of Biochemistry, Germany) for providing the HA-ubiquitin construct and Peter Becker (Adolf Butenandt Institute, Germany) for the 601 DNA

construct. We are grateful to the following colleagues for providing ESCs and somatic cells: Masahiro Muto and Haruhiko Koseki for mouse E14 wt and *Uhrfl*^{-/-} ESCs; En Li and T. Chen for mouse J1 wt and *Dnmt1*^{-/-}; Thomas Jenuwein for MEF cells; and L. David Spector for providing BHK cells containing a lac operator repeat array. We thank E.M. Baur (Ludwig Maximilians University, Germany) for technical help with the GFP-TS UIM point mutant plasmid constructs. This work was supported by grants from the Deutsche Forschungsgemeinschaft (DFG, SFB1064 A17 to HL and Z03 to AI), the Nanosystem Initiative Munich (NIM to HL) and the Epigenomics Flagship Project (EPIGEN-CNR -IT to IMB). KF and GP were supported by the International Max Planck Research School for Molecular and Cellular Life Sciences (IM-PRS-LS). PW, NL and MS are fellows of the Graduate School Life Science Munich (LSM). MS is a fellow of the Integrated Research Training Group (IRTG) of the SFB1064. NL and WQ were also supported by the China Scholarship Council (CSC).

References

- Bird A. DNA methylation patterns and epigenetic memory. *Genes Dev* 2002; **16**:6-21.
- Leonhardt H, Page AW, Weier HU, Bestor TH. A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. *Cell* 1992; **71**:865-873.
- Chuang LS, Ian HI, Koh TW, Ng HH, Xu G, Li BF. Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. *Science* 1997; **277**:1996-2000.
- Schneider K, Fuchs C, Dobay A, et al. Dissection of cell cycle-dependent dynamics of Dnmt1 by FRAP and diffusion-coupled modeling. *Nucleic Acids Res* 2013; **41**:4860-4876.
- Easwaran HP, Schermelleh L, Leonhardt H, Cardoso MC. Replication-independent chromatin loading of Dnmt1 during G2 and M phases. *EMBO Rep* 2004; **5**:1181-1186.
- Schermelleh L, Haemmer A, Spada F, et al. Dynamics of Dnmt1 interaction with the replication machinery and its role in postreplicative maintenance of DNA methylation. *Nucleic Acids Res* 2007; **35**:4301-4312.
- Spada F, Haemmer A, Kuch D, et al. DNMT1 but not its interaction with the replication machinery is required for maintenance of DNA methylation in human cells. *J Cell Biol* 2007; **176**:565-571.
- Margot JB, Aguirre-Arteta AM, Di Giacco BV, et al. Structure and function of the mouse DNA methyltransferase gene: Dnmt1 shows a tripartite structure. *J Mol Biol* 2000; **297**:293-300.
- Fellinger K, Rothbauer U, Felle M, Langst G, Leonhardt H. Dimerization of DNA methyltransferase 1 is mediated by its regulatory domain. *J Cell Biochem* 2009; **106**:521-528.
- Syeda F, Fagan RL, Wean M, et al. The replication focus targeting sequence (RFTS) domain is a DNA-competitive inhibitor of Dnmt1. *J Biol Chem* 2011; **286**:15344-15351.
- Takeshita K, Suetake I, Yamashita E, et al. Structural insight into maintenance methylation by mouse DNA methyltransferase 1 (Dnmt1). *Proc Natl Acad Sci USA* 2011; **108**:9055-9059.
- Achour M, Jacq X, Ronde P, et al. The interaction of the SRA domain of ICBP90 with a novel domain of DNMT1 is involved in the regulation of VEGF gene expression. *Oncogene*

- 2008; **27**:2187-2197.
- 13 Felle M, Joppien S, Nemeth A, *et al.* The USP7/Dnmt1 complex stimulates the DNA methylation activity of Dnmt1 and regulates the stability of UHRF1. *Nucleic Acids Res* 2011; **39**:8355-8365.
- 14 Berkkyurek AC, Suetake I, Arita K, *et al.* The DNA Methyltransferase Dnmt1 directly interacts with the SET and RING finger associated (SRA) domain of the multifunctional protein Uhrf1 to facilitate accession of the catalytic center to hemi-methylated DNA. *J Biol Chem* 2013; **289**:379-386.
- 15 Zhang J, Gao Q, Li P, *et al.* S phase-dependent interaction with DNMT1 dictates the role of UHRF1 but not UHRF2 in DNA methylation maintenance. *Cell Res* 2011; **21**:1723-1739.
- 16 Bostick M, Kim JK, Esteve PO, Clark A, Pradhan S, Jacobsen SE. UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science* 2007; **317**:1760-1764.
- 17 Sharif J, Muto M, Takebayashi S, *et al.* The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature* 2007; **450**:908-912.
- 18 Li E, Bestor TH, Jaenisch R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* 1992; **69**:915-926.
- 19 Arita K, Ariyoshi M, Tochio H, Nakamura Y, Shirakawa M. Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. *Nature* 2008; **455**:818-821.
- 20 Avvakumov GV, Walker JR, Xue S, *et al.* Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. *Nature* 2008; **455**:822-825.
- 21 Qian C, Li S, Jakoncic J, Zeng L, Walsh MJ, Zhou MM. Structure and hemimethylated CpG binding of the SRA domain from human UHRF1. *J Biol Chem* 2008; **283**:34490-34494.
- 22 Rottach A, Frauer C, Pichler G, Bonapace IM, Spada F, Leonhardt H. The multi-domain protein Np95 connects DNA methylation and histone modification. *Nucleic Acids Res* 2010; **38**:1796-1804.
- 23 Cheng J, Yang Y, Fang J, *et al.* Structural insight into coordinated recognition of trimethylated histone H3 lysine 9 (H3K9me3) by the plant homeodomain (PHD) and tandem tudor domain (TTD) of UHRF1 (ubiquitin-like, containing PHD and RING finger domains, 1) protein. *J Biol Chem* 2013; **288**:1329-1339.
- 24 Citterio E, Papait R, Nicassio F, *et al.* Np95 is a histone-binding protein endowed with ubiquitin ligase activity. *Mol Cell Biol* 2004; **24**:2526-2535.
- 25 Karagianni P, Amazit L, Qin J, Wong J. ICBP90, a novel methyl K9 H3 binding protein linking protein ubiquitination with heterochromatin formation. *Mol Cell Biol* 2008; **28**:705-717.
- 26 Xie S, Jakoncic J, Qian C. UHRF1 double tudor domain and the adjacent PHD finger act together to recognize K9me3-containing histone H3 tail. *J Mol Biol* 2012; **415**:318-328.
- 27 Papait R, Pistore C, Grazini U, *et al.* The PHD domain of Np95 (mUHRF1) is involved in large-scale reorganization of pericentromeric heterochromatin. *Mol Biol Cell* 2008; **19**:3554-3563.
- 28 Jenkins Y, Markovtsov V, Lang W, *et al.* Critical role of the ubiquitin ligase activity of UHRF1, a nuclear RING finger protein, in tumor cell growth. *Mol Biol Cell* 2005; **16**:5621-5629.
- 29 Qin W, Leonhardt H, Spada F. Usp7 and Uhrf1 control ubiquitination and stability of the maintenance DNA methyltransferase Dnmt1. *J Cell Biochem* 2011; **112**:439-444.
- 30 Du Z, Song J, Wang Y, *et al.* DNMT1 stability is regulated by proteins coordinating deubiquitination and acetylation-driven ubiquitination. *Sci Signal* 2010; **3**:ra80.
- 31 Mudbhary R, Hoshida Y, Chernyavskaya Y, *et al.* UHRF1 overexpression drives DNA hypomethylation and hepatocellular carcinoma. *Cancer Cell* 2014; **25**:196-209.
- 32 Nishiyama A, Yamaguchi L, Sharif J, *et al.* Uhrf1-dependent H3K23 ubiquitylation couples maintenance DNA methylation and replication. *Nature* 2013; **502**:249-253.
- 33 Schermelleh L, Spada F, Easwaran HP, *et al.* Trapped in action: direct visualization of DNA methyltransferase activity in living cells. *Nat Methods* 2005; **2**:751-756.
- 34 Liu X, Gao Q, Li P, *et al.* UHRF1 targets DNMT1 for DNA methylation through cooperative binding of hemi-methylated DNA and methylated H3K9. *Nat Commun* 2013; **4**:1563.
- 35 LDambacher S, Deng W, Hahn M, *et al.* CENP-C facilitates the recruitment of M18BP1 to centromeric chromatin. *Nucleus* 2012; **3**:101-110.
- 36 Zolghadr K, Mortusewicz O, Rothbauer U, *et al.* A fluorescent two-hybrid assay for direct visualization of protein interactions in living cells. *Mol Cell Proteomics* 2008; **7**:2279-2287.
- 37 Hu L, Li Z, Wang P, Lin Y, Xu Y. Crystal structure of PHD domain of UHRF1 and insights into recognition of unmodified histone H3 arginine residue 2. *Cell Res* 2011; **21**:1374-1378.
- 38 Rajakumara E, Wang Z, Ma H, *et al.* PHD finger recognition of unmodified histone H3R2 links UHRF1 to regulation of euchromatic gene expression. *Mol Cell* 2011; **43**:275-284.
- 39 Wang C, Shen J, Yang Z *et al.* Structural basis for site-specific reading of unmodified R2 of histone H3 tail by UHRF1 PHD finger. *Cell Res* 2011; **21**:1379-1382.
- 40 Frauer C, Leonhardt H. A versatile non-radioactive assay for DNA methyltransferase activity and DNA binding. *Nucleic Acids Res* 2009; **37**:e22.
- 41 Bestor TH, Ingram VM. Two DNA methyltransferases from murine erythroleukemia cells: purification, sequence specificity, and mode of interaction with DNA. *Proc Natl Acad Sci USA* 1983; **80**:5559-5563.
- 42 Yoder JA, Soman NS, Verdine GL, Bestor TH. DNA (cytosine-5)-methyltransferases in mouse cells and tissues. Studies with a mechanism-based probe. *J Mol Biol* 1997; **270**:385-395.
- 43 Jeltsch A. On the enzymatic properties of Dnmt1: specificity, processivity, mechanism of linear diffusion and allosteric regulation of the enzyme. *Epigenetics* 2006; **1**:63-66.
- 44 Bashtrykov P, Jankevicius G, Jurkowska RZ, Ragozin S, Jeltsch A. The UHRF1 protein stimulates the activity and specificity of the maintenance DNA methyltransferase DNMT1 by an allosteric mechanism. *J Biol Chem* 2014; **289**:4106-4115.
- 45 Rothbart SB, Dickson BM, Ong MS, *et al.* Multivalent histone engagement by the linked tandem Tudor and PHD domains of UHRF1 is required for the epigenetic inheritance of DNA methylation. *Genes Dev* 2013; **27**:1288-1298.
- 46 Rothbart SB, Krajewski K, Nady N, *et al.* Association of UHRF1 with methylated H3K9 directs the maintenance of DNA methylation. *Nat Struct Mol Biol* 2012; **19**:1155-1160.
- 47 Song J, Rechkoblit O, Bestor TH, Patel DJ. Structure of DNMT1-DNA complex reveals a role for autoinhibition in main-

- tenance DNA methylation. *Science* 2011; **331**:1036-1040.
- 48 Kalb R, Latwiel S, Baymaz HI, et al. Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression. *Nat Struct Mol Biol* 2014; **21**:569-571.
- 49 Leeb M, Wutz A. RING1B is crucial for the regulation of developmental control genes and PRC1 proteins but not X inactivation in embryonic cells. *J Cell Biol* 2007; **178**:219-229.
- 50 Blackledge NP, Farcas AM, Kondo T, et al. Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation. *Cell* 2014; **157**:1445-1459.
- 51 Vire E, Brenner C, Deplus R, et al. The Polycomb group protein EZH2 directly controls DNA methylation. *Nature* 2006; **439**:871-874.
- 52 Cooper S, Dienstbier M, Hassan R, et al. Targeting polycomb to pericentric heterochromatin in embryonic stem cells reveals a role for H2AK119u1 in PRC2 recruitment. *Cell reports* 2014; **7**:1456-1470.
- 53 Kurdistan SK, Tavazoie S, Grunstein M. Mapping global histone acetylation patterns to gene expression. *Cell* 2004; **117**:721-733.
- 54 Tsai WW, Wang Z, Yiu TT, et al. TRIM24 links a non-canonical histone signature to breast cancer. *Nature* 2010; **468**:927-932.
- 55 Wang Z, Zang C, Rosenfeld JA, et al. Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat Genet* 2008; **40**:897-903.
- 56 Raiborg C, Stenmark H. The ESCRT machinery in endosomal sorting of ubiquitylated membrane proteins. *Nature* 2009; **458**:445-452.
- 57 Lim J, Son WS, Park JK, Kim EE, Lee BJ, Ahn HC. Solution structure of UIM and interaction of tandem ubiquitin binding domains in STAM1 with ubiquitin. *Biochem Biophys Res Commun* 2011; **405**:24-30.
- 58 Hirano S, Kawasaki M, Ura H, et al. Double-sided ubiquitin binding of Hrs-UIM in endosomal protein sorting. *Nat Struct Mol Biol* 2006; **13**:272-277.
- 59 Wang Q, Young P, Walters KJ. Structure of S5a bound to monoubiquitin provides a model for polyubiquitin recognition. *J Mol Biol* 2005; **348**:727-739.
- 60 Frauer C, Rottach A, Meilinger D, et al. Different binding properties and function of CXXC zinc finger domains in Dnmt1 and Tet1. *PLoS One* 2011; **6**:e16627.
- 61 Meilinger D, Fellingner K, Bultmann S, et al. Np95 interacts with de novo DNA methyltransferases, Dnmt3a and Dnmt3b, and mediates epigenetic silencing of the viral CMV promoter in embryonic stem cells. *EMBO Rep* 2009; **10**:1259-1264.
- 62 Ho SN, Hunt HD, Horton RM, Pullen JK, Pease LR. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* 1989; **77**:51-59.
- 63 Pichler G, Wolf P, Schmidt CS, et al. Cooperative DNA and histone binding by Uhrf2 links the two major repressive epigenetic pathways. *J Cell Biochem* 2011; **112**:2585-2593.
- 64 Rothbauer U, Zolghadr K, Tillib S, et al. Targeting and tracing antigens in live cells with fluorescent nanobodies. *Nat Methods* 2006; **3**:887-889.
- 65 Herce HD, Deng W, Helma J, Leonhardt H, Cardoso MC. Visualization and targeted disruption of protein interactions in living cells. *Nat Commun* 2013; **4**:2660.
- 66 Rottach A, Kremmer E, Nowak D, et al. Generation and characterization of a rat monoclonal antibody specific for PCNA. *Hybridoma (Larchmt)* 2008; **27**:91-98.
- 67 Rottach A, Kremmer E, Nowak D, Leonhardt H, Cardoso MC. Generation and characterization of a rat monoclonal antibody specific for multiple red fluorescent proteins. *Hybridoma (Larchmt)* 2008; **27**:337-343.
- 68 Lei H, Oh SP, Okano M, et al. De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. *Development* 1996; **122**:3195-3205.
- 69 Tucker KL, Beard C, Dausmann J, et al. Germ-line passage is required for establishment of methylation and expression patterns of imprinted but not of nonimprinted genes. *Genes Dev* 1996; **10**:1008-1020.
- 70 Tost J, Gut IG. DNA methylation analysis by pyrosequencing. *Nat Protoc* 2007; **2**:2265-2275.
- 71 Shechter D, Dormann HL, Allis CD, Hake SB. Extraction, purification and analysis of histones. *Nat Protoc* 2007; **2**:1445-1457.
- 72 Wysocka J. Identifying novel proteins recognizing histone modifications using peptide pull-down assay. *Methods* 2006; **40**:339-343.
- 73 Villar-Garea A, Israel L, Imhof A. Analysis of histone modifications by mass spectrometry. *Curr Protoc Protein Sci* 2008; Chapter 14:Unit 14.10.
- 74 Olsen JV, de Godoy LM, Li G, et al. Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. *Mol Cell Proteomics* 2005; **4**:2010-2021.
- 75 Mark Hall EF, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten. The WEKA data mining software: an update. *ACM SIGKDD Explorations Newsletter* 2009; **11**:10-18.
- 76 Schindelin J, Arganda-Carreras I, Frise E, et al. Fiji: an open-source platform for biological-image analysis. *Nat Methods* 2012; **9**:676-682.
- 77 R Core Team (2014). R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- 78 Robert X, Gouet P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* 2014; **42**:W320-W324.

(Supplementary information is linked to the online version of the paper on the *Cell Research* website.)



This work is licensed under the Creative Commons Attribution-NonCommercial-No Derivative Works 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0>

Supplementary information, Data S1

Materials and methods

Antibodies

For detection of (GFP)-DNMT1 by western blot, a mouse anti-DNMT1 antibody (pATH52 [1, 2]) was used. Equal loading was confirmed by immunoblotting with a specific anti-Lamin B1 (Abcam) or anti- β -Actin (Sigma) antibody. The rabbit anti-H2A antibody was purchased from Millipore.

In vitro DNA methylation assay

For analyses of *in vitro* DNA methylation activity, GFP-DNMT1 was purified by immunoprecipitation from HEK 293T extracts. The concentration of GFP-DNMT1 in the bound fractions was measured by fluorescent read out. In order to get enough unmodified DNA templates, the 601 DNA sequences were amplified with the primers: TGCATGTATTGAACAG (forward) and TGCACAGGATGTATATATC (reverse). 3 μ g of GFP-DNMT1 were incubated with 88 ng of DNA template in methylation buffer containing 160 μ M SAM and 100 ng/ μ l BSA at 37°C for 3 hours. After inactivation of the reaction at 65°C for 30 min, the DNA was isolated with a Nucleospin PCR cleaning kit (Macherey-Nagel) and bisulfite treated with EZ DNA Methylation-Gold Kit (Zymo research). Primer sequences for the 601 DNA were TGTATGTATTGAATAG (forward primer) and TACACAAAATATATATATC (reverse primer). For amplification we used Qiagen Hot Start Polymerase in 1x Qiagen Hot Start Polymerase buffer supplemented with 0.2 mM dNTPs, 0.2 μ M forward primer, 0.2 μ M reverse primer, 1.3 mM Betaine (Sigma) and 60 mM Tetramethylammonium-chloride (TMAC, Sigma). Pyrosequencing reactions were carried out by Varionostic GmbH (Ulm).

Preparation of hemimethylated DNA substrates

To prepare the hemimethylated DNA, an efficient method for long heteroduplex DNA was used as described [3]. One pair of modified PCR primers were synthesized, which are labeled with phosphate at 5'-end, 5'-phosphorylated-TGCATGTATTGAACAG-3' and 5'-phosphorylated-TGCACAGGATGTATATATC-3'. To get single and upper strand DNA, the DNA was amplified with the reverse primer labeled with phosphate at the 5'-end, following a lambda-nuclease digestion (NEB). The same procedure is required for making lower strand DNA. To prepare the methylated lower strand DNA, one more step, *in vitro* methylation by bacterial methyltransferase M.SssI (NEB), is required before treatment with the lambda-nuclease. In the end, equal amounts of upper and lower strand DNA were mixed and incubated at 95°C for 5 min, followed by annealing. To get rid of contamination from double strand DNA after lambda-nuclease treatment, the hydroxyapatite chromatography was carried out. Hydroxyapatite column (Sigma) was packed according to the manufacturer's instructions and the single stranded DNA was eluted by elution buffer containing 150 mM sodium phosphate.

In vitro DNA binding assays

In vitro DNA binding assays were performed as described previously [4]. Briefly, two double stranded DNA oligonucleotides labeled with different ATTO fluorophores were used as substrates in direct competition. DNA oligonucleotide substrates with identical sequence contained an unmodified or hemimethylated cytosine at a single, central CpG site (UMB: unmodified binding substrate, ATTO550; HMB: hemimethylated binding substrates, ATTO647N; Supplementary Table S3). GFP fusion proteins were expressed in HEK 293T cells and immunoprecipitated using the GFP-Trap (Chromotek). Immobilized UHRF1-GFP wt and mutants were washed three times before incubation with DNA substrates at a final concentration of 160 nM each. After removal of unbound substrates, protein amounts (GFP fluorescence) and bound DNA were measured with an Infinite M1000 plate reader (Tecan).

In vivo autoubiquitination assay

The *in vivo* autoubiquitination assay of UHRF1-GFP was performed as described before [5]. The resulting ubiquitination levels were detected with a specific mouse monoclonal anti-HA antibody (12CA5) and quantified using Image J and a statistical Student's t-test analysis. Equal amounts of (UHRF1-)GFP in the bound fraction were verified by immunoblotting with a specific anti-GFP antibody (Chromotek).

Slot blot analysis

To quantify global DNA methylation levels, the Bio-Rad slot blot system was used according to the manufacturer's instruction. Prior to loading on a Nitrocellulose membrane (Amersham), genomic DNA was denatured in 6x SSC buffer for 10 min at 95°C and incubated for 10 min on ice. The membrane was crosslinked, blocked with 5% milk and immunostained with specific rabbit anti-ssDNA (Eurogentec) and mouse anti-5mC (IBL) antibodies. Quantification was performed using the ImageJ gel analysis tool.

In vitro histone tail peptide binding assay

The *in vitro* histone tail peptide binding assay was performed as described before [6] with the following modification. GFP fusion proteins were equalized to a GFP concentration of 130 nM prior to immunoprecipitation with the GFP-Trap. The TAMRA-labeled H3 peptides used in this assay are listed in Supplementary Table S2.

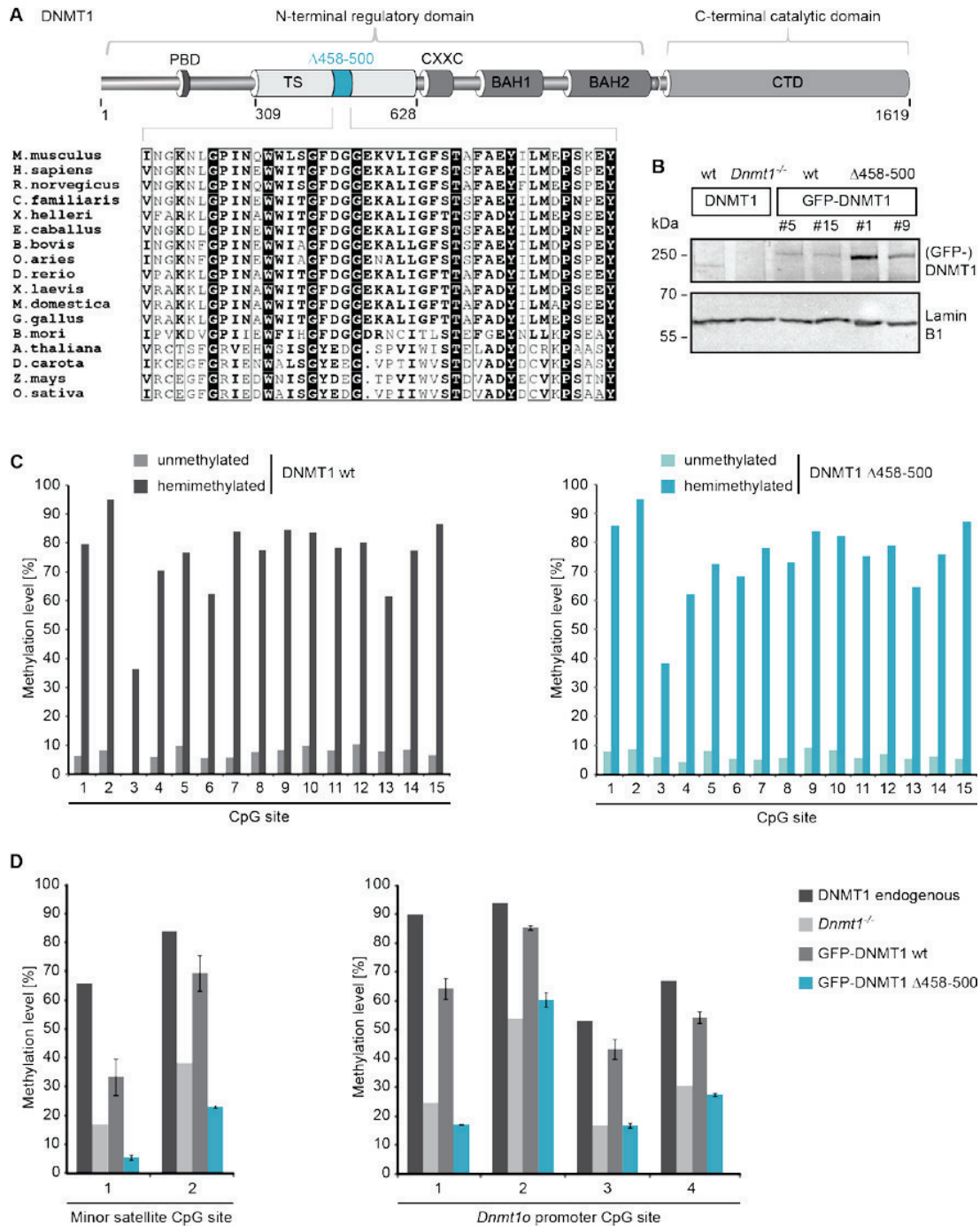
Ubiquitinated histone H3 binding experiment

For ubiquitinated H3 binding experiment, HEK 293T cells were incubated with 2 mM N-Ethylmaleimide (NEM, AppliChem) for 10 min before harvesting and were treated in hypotonic buffer (10 mM Tris-HCl pH 8, 10 mM KCl, 1.5 mM MgCl₂, 1 mM DTT and 1x Protease Inhibitor, 2 mM PMSF) for 10 min on ice to isolate the intact nuclei. Nuclei resuspended in MNase digestion buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM CaCl₂, 0.1% NP-40, 1x Protease Inhibitor (Serva), 2 mM PMSF) were digested with 40 U/ml MNase at 37°C for 5 min to get mononucleosomes.

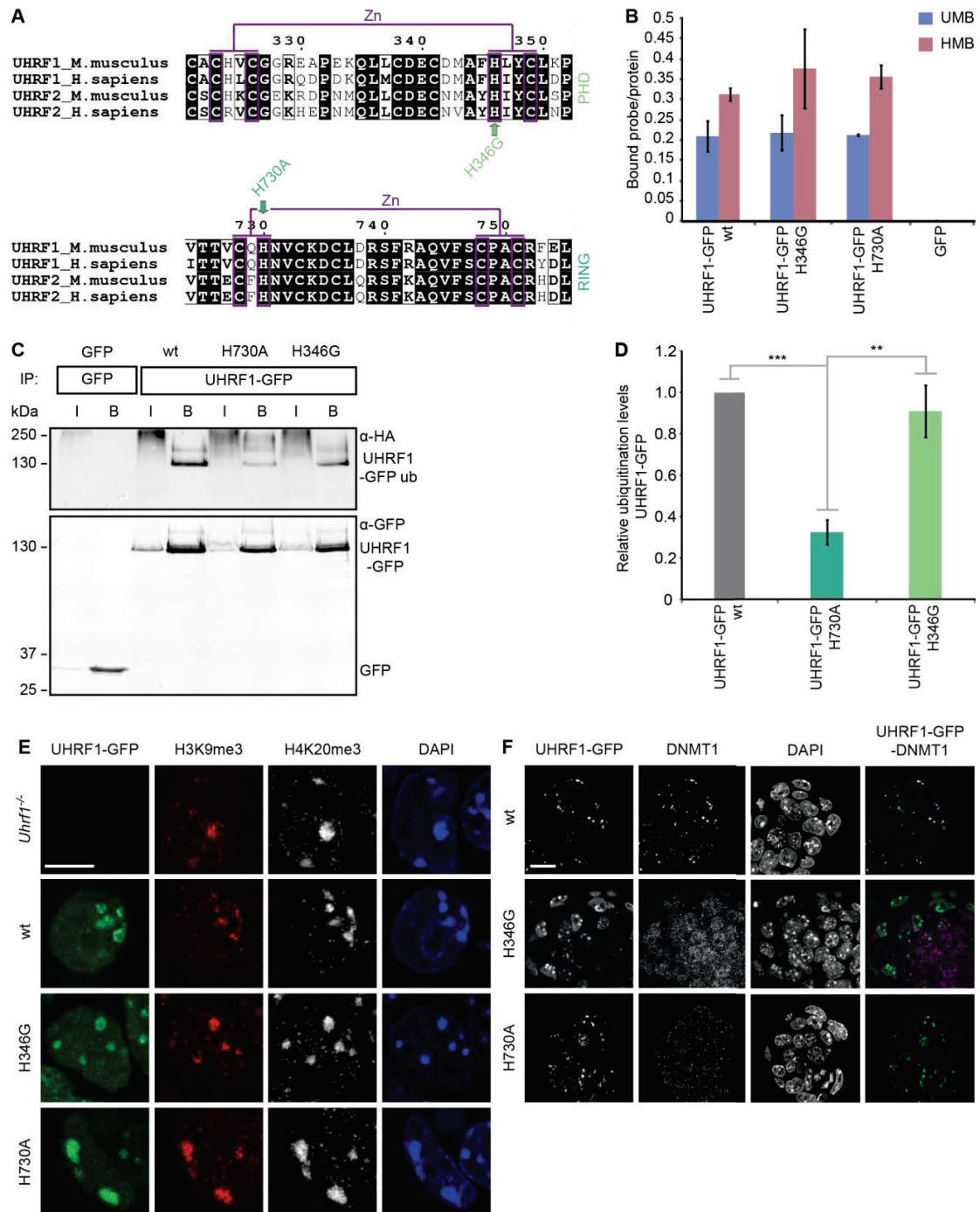
GFP-DNMT1 and its mutants were immobilized on the GFP-Trap and incubated with equal amount of mononucleosomes for 2 hours. After washing steps, the bound fractions were analyzed by western blot.

Supplementary references

- 1 Bestor TH. Activation of mammalian DNA methyltransferase by cleavage of a Zn binding regulatory domain. *EMBO J* 1992; **11**:2611-2617.
- 2 Li E, Bestor TH, Jaenisch R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* 1992; **69**:915-926.
- 3 Thomas E, Pingoud A, Friedhoff P. An efficient method for the preparation of long heteroduplex DNA as substrate for mismatch repair by the Escherichia coli MutHLS system. *Biol Chem* 2002; **383**:1459-1462.
- 4 Frauer C, Leonhardt H. A versatile non-radioactive assay for DNA methyltransferase activity and DNA binding. *Nucleic Acids Res* 2009; **37**:e22.
- 5 Qin W, Leonhardt H, Spada F. Usp7 and Uhrf1 control ubiquitination and stability of the maintenance DNA methyltransferase Dnmt1. *J Cell Biochem* 2011; **112**:439-444.
- 6 Pichler G, Wolf P, Schmidt CS et al. Cooperative DNA and histone binding by Uhrf2 links the two major repressive epigenetic pathways. *J Cell Biochem* 2011; **112**:2585-2593.
- 7 Robert X, Gouet P. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* 2014; **42**:W320-324.
- 8 Hirano S, Kawasaki M, Ura H et al. Double-sided ubiquitin binding of Hrs-UIP in endosomal protein sorting. *Nat Struct Mol Biol* 2006; **13**:272-277.

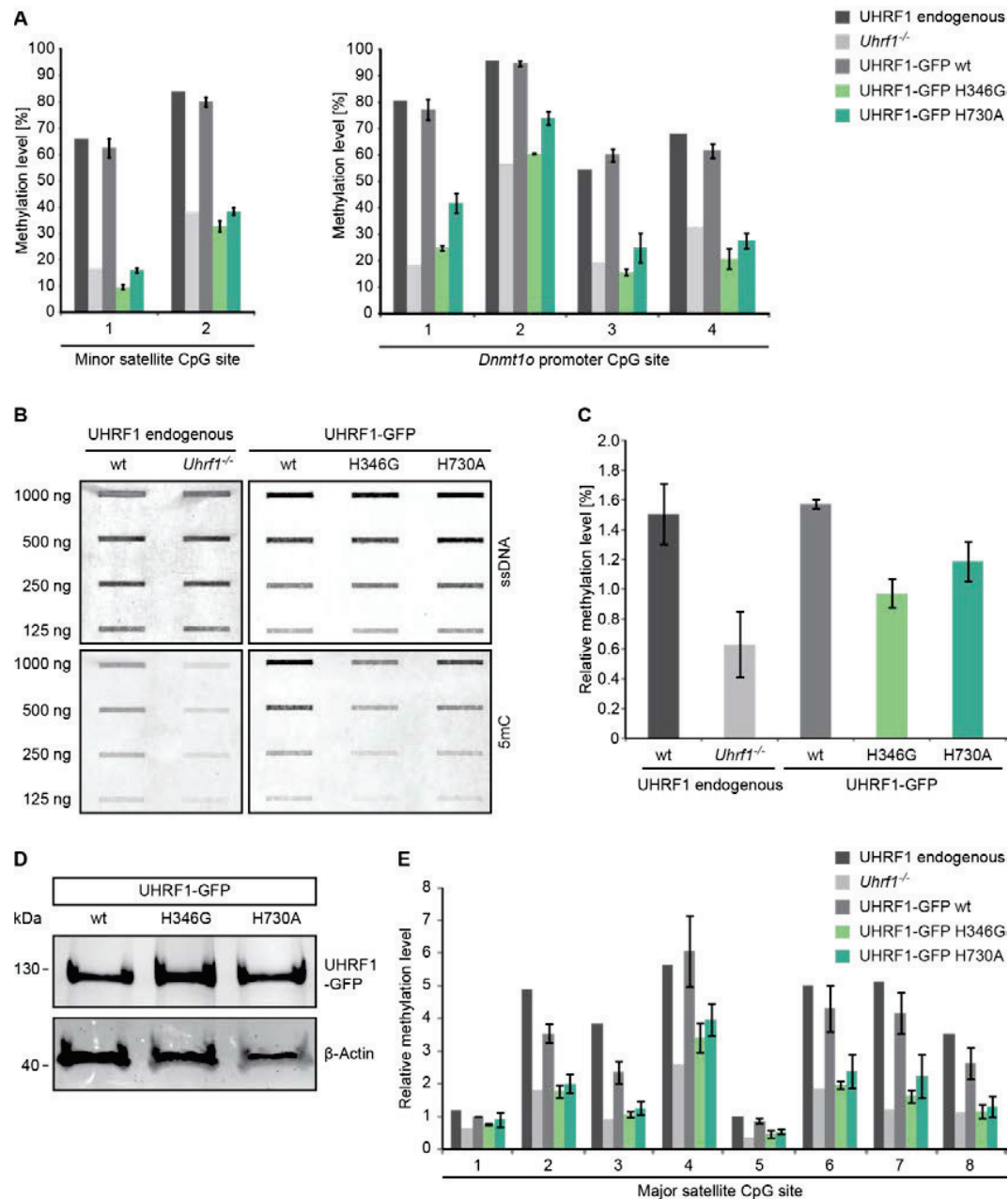


Supplementary information, Figure S1 GFP-DNMT1 carrying a deletion of amino acids 458-500 in the TS domain is still active *in vitro*, but inactive *in vivo*. (A) Schematic outline of DNMT1 domains and the TS domain deletion (Δ458-500). The deleted region comprises a conserved core of the TS domain showing high sequence similarities among higher eukaryotes. The alignment was generated using ESPrnt [7]. (B) Expression levels of GFP-DNMT1 wt (clones #5 and #15) and the TS domain deletion mutant (Δ458-500, clones #1 and #9) in stable ESC lines were compared to the endogenous DNMT1 level in J1 wt cells. Lamin B1 is shown as loading control. DNMT1 was detected with a specific anti-DNMT1 antibody. (C) *In vitro* methylation assay of GFP-DNMT1 wt and Δ458-500 mutant using unmodified or hemimethylated DNA as a substrate. (D) Local DNA methylation analyses at minor satellite repeats and the *Dnmt1o* promoter. CpG methylation levels of mouse J1 *Dnmt1*^{-/-} ESCs stably expressing GFP-DNMT1 wt or Δ458-500 mutant constructs were analyzed by bisulfite treatment of genomic DNA, PCR amplification and direct pyrosequencing. The methylation level of the J1 wt cell line (endogenous DNMT1) and untransfected J1 *Dnmt1*^{-/-} cells are shown for comparison. Displayed are mean values ± SD from two different clones.

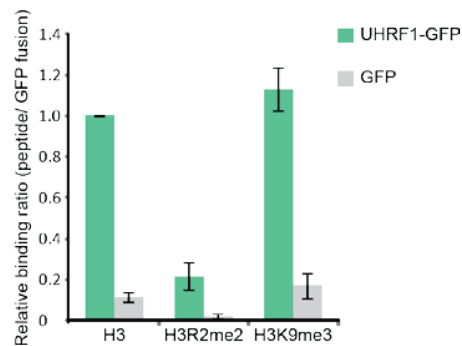


Supplementary information, Figure S2 The UHRF1-GFP PHD and RING domain mutants preserve their preference for hemimethylated DNA and their heterochromatin association, but cannot recruit endogenous DNMT1 to heterochromatin. **(A)** Primary sequence alignment of mouse and human UHRF1 and UHRF2 at regions of the PHD and RING domain, respectively. Note that the mutations affect conserved amino acids contributing to the coordination of zinc (Zn) ions. **(B)** *In vitro* DNA binding assay. The binding of immunoprecipitated UHRF1-GFP wt or PHD (H346G) and RING domain (H730A) mutants to either un- (UMB) or hemimethylated (HMB) fluorescently labeled double stranded oligonucleotide probes was tested in direct competition. Shown are mean fluorescence intensity ratios of bound probe over bound GFP fusion of three independent experiments \pm SD. GFP was used as negative control. **(C)** *In vivo* autoubiquitination assay of UHRF1-GFP. Wt or PHD and RING domain mutant constructs were transiently co-expressed with HA-ubiquitin in HEK 293T cells and UHRF1-GFP was immunoprecipitated using the GFP-Trap. Ubiquitination levels were detected by immunoblotting with an anti-HA antibody and bound fractions were verified with a specific anti-GFP antibody. GFP was used as negative control. I, input; B, bound. One representative blot of three independent replicates is depicted. **(D)** Quantification of the *in vivo* autoubiquitination activity of UHRF1-GFP wt or PHD and RING domain mutant

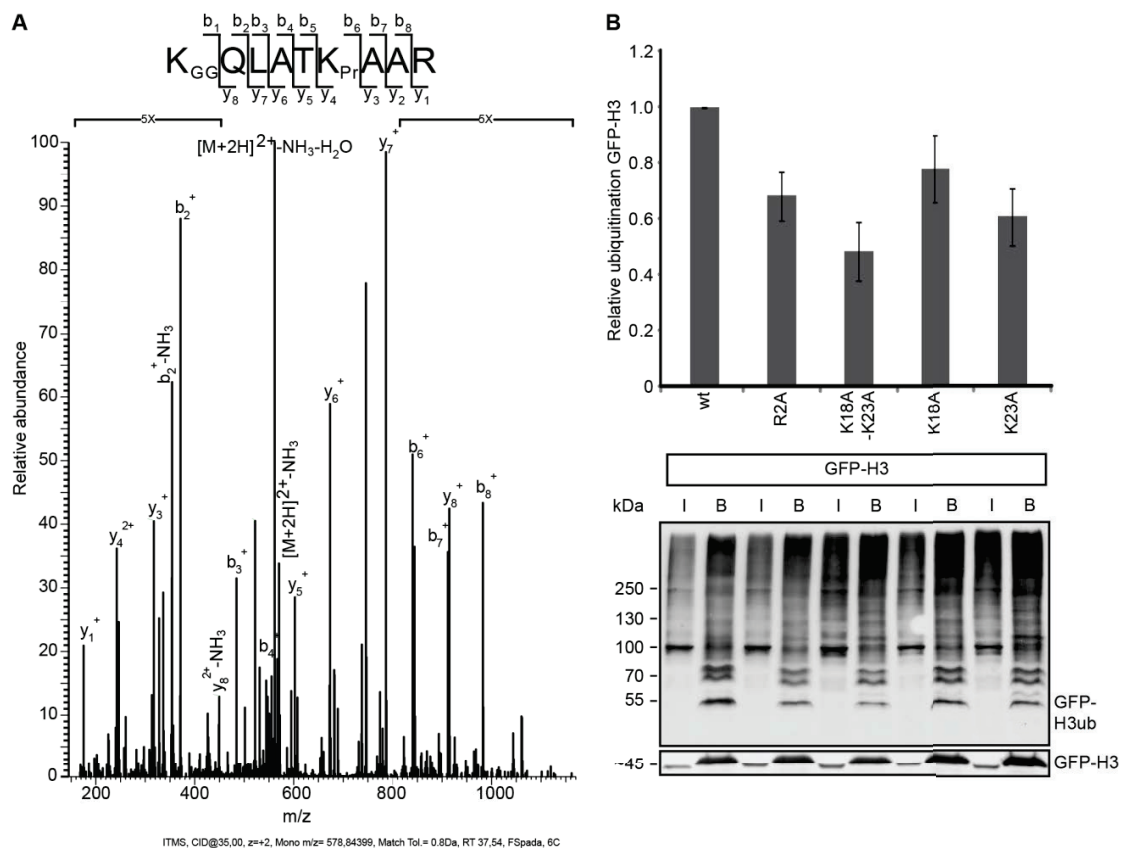
constructs depicted in (C). Shown are mean values \pm SD of three independent biological replicates analyzed using ImageJ and normalized to the ubiquitination level of the UHRF1-GFP wt construct. Differences between the UHRF1-GFP wt or PHD mutant and the RING domain mutant were analyzed using a Student's t-test and considered statistically significant for $P < 0.01$ (**) and highly significant for $P < 0.001$ (***). **(E)** and **(F)** Confocal mid sections of fixed E14 *Uhrf1*^{-/-} ESCs stably expressing UHRF1-GFP wt or PHD (H346G) and RING domain (H730A) mutant constructs. **(E)** H3K9me3 and H4K20me3 were stained with specific antibodies as markers for heterochromatin and DNA was counterstained with DAPI. Scale bar, 5 μ m. **(F)** Endogenous DNMT1 in ESCs stably expressing UHRF1-GFP wt or its PHD and RING domain mutants was stained by a specific antibody and DNA was counterstained with DAPI. Scale bar, 10 μ m.



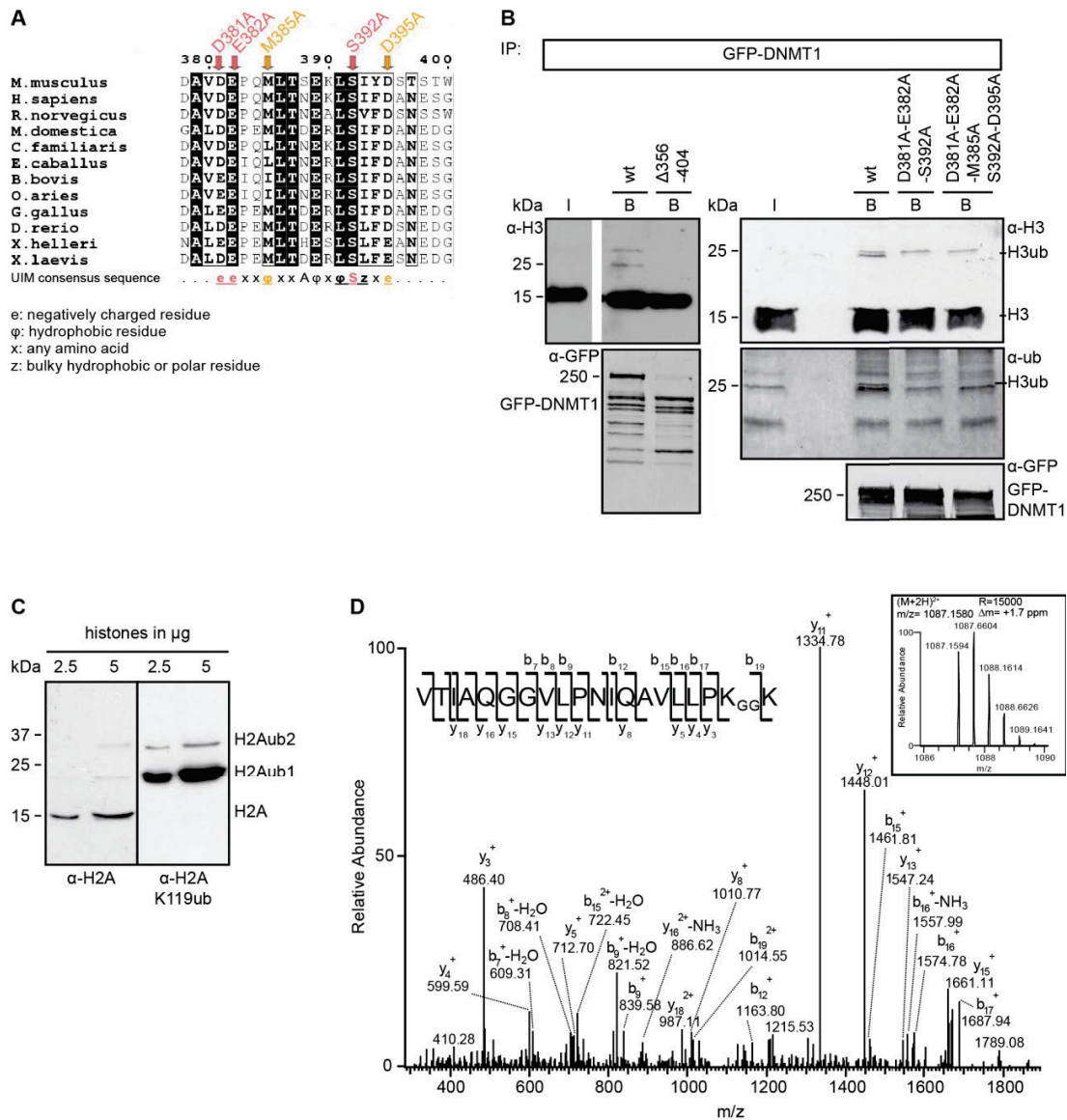
Supplementary information, Figure S3 The UHRF1-GFP PHD and RING domain mutants cannot mediate DNA remethylation in mouse E14 *Uhrf1*^{-/-} ESCs. **(A)** Local DNA methylation analyses at minor satellite repeats and the *Dnmt1* promoter. CpG methylation levels of mouse E14 *Uhrf1*^{-/-} ESCs stably expressing UHRF1-GFP wt or PHD (H346G) and RING domain (H730A) mutant constructs were analyzed by bisulfite treatment of genomic DNA, PCR amplification and direct pyrosequencing. The methylation level of E14 wt ESCs (endogenous UHRF1) and untransfected E14 *Uhrf1*^{-/-} cells are shown for comparison. Mean values \pm SD from two different clones were calculated. **(B)** and **(C)** Global DNA methylation analysis of stable UHRF1-GFP wt or PHD and RING domain mutant cell lines compared to E14 wt ESCs expressing endogenous UHRF1 and E14 *Uhrf1*^{-/-} ESCs. **(B)** Slot blot analysis using a dilution series of genomic DNA and immunodetection with specific anti-single stranded (ss) DNA and anti-5-methylcytosine (5mC) antibodies. A representative slot blot of technical duplicates is shown. **(C)** The 5mC signals relative to the signals of ssDNA were quantified using the ImageJ gel analysis tool. Relative global DNA methylation levels are indicated as mean values \pm SD of technical duplicates. **(D)** Expression analysis of UHRF1-GFP wt or PHD and RING domain mutants after transient transfection in E14 *Uhrf1*^{-/-} ESCs and FACS sorting. Fusion proteins were detected with a specific anti-GFP antibody and β -Actin serves as a loading control. **(E)** Local DNA methylation analyses of E14 *Uhrf1*^{-/-} ESCs transiently expressing GFP-UHRF1 wt or PHD (H346G) and RING domain (H730A) mutant constructs. CpG methylation levels at the major satellite repeats were analyzed by bisulfite treatment of genomic DNA, PCR amplification and direct pyrosequencing. Shown are mean values \pm SD from three independent experiments normalized to the value of UHRF1-GFP wt at the first CpG site.



Supplementary information, Figure S4 Histone H3 binding by UHRF1-GFP is sensitive to R2 methylation. Fluorescent *in vitro* histone tail peptide binding assay of UHRF1-GFP using TAMRA labeled H3 tails carrying no modification (H3), an asymmetrically dimethylated arginine 2 (H3R2me2) or a trimethylated lysine 9 (H3K9me3). Shown are mean values of the relative binding ratio of histone tail peptides over GFP fusion proteins from five independent experiments \pm SEM. GFP was used as negative control.

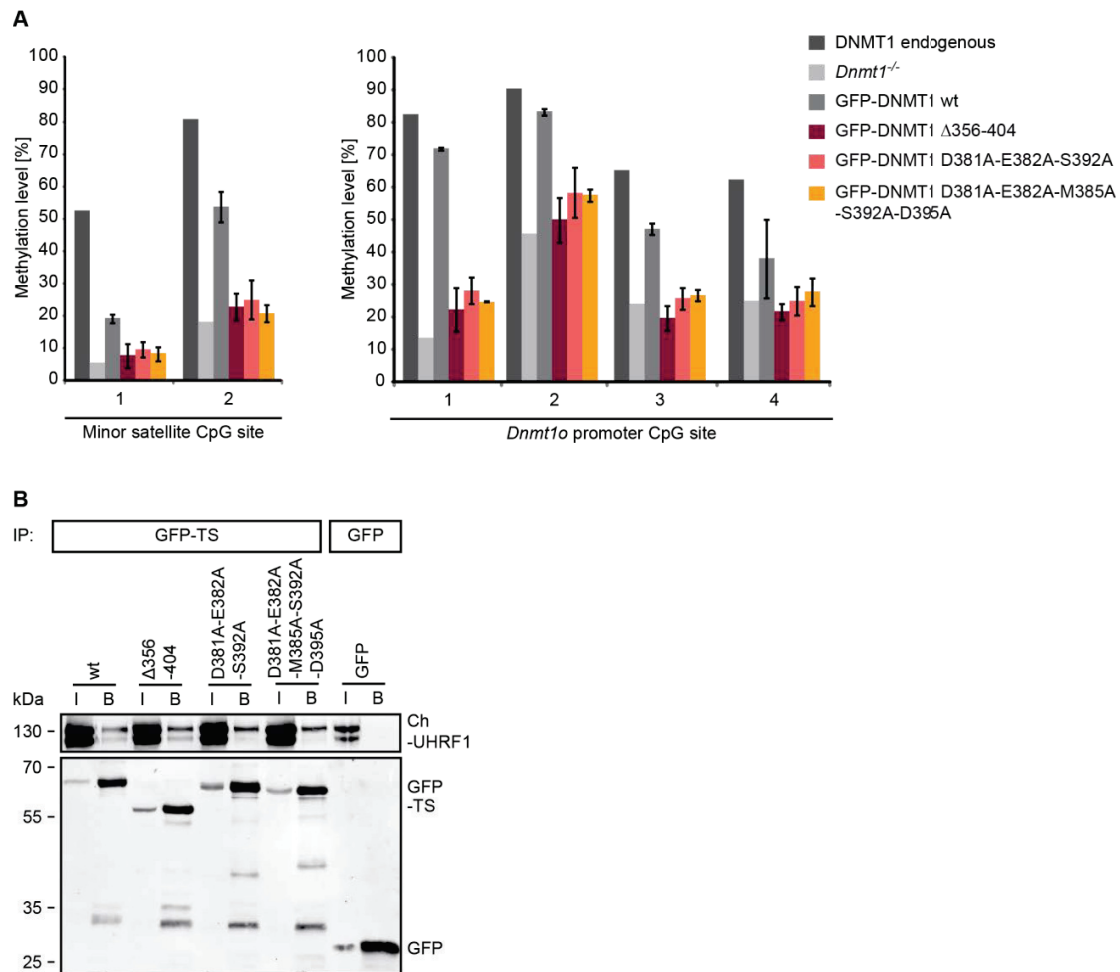


Supplementary information, Figure S5 H3 histone tail ubiquitination and dependence on R2. **(A)** After expression of UHRF1-GFP and HA-Ubiquitin in HEK 293T cells, acid extracted ubiquitinated histones were immunoprecipitated with an anti-HA antibody and analyzed by mass spectrometry. The Collision-induced dissociation (CID) MS/MS spectrum of the histone H3 18-26 peptide shows ubiquitination (GG) at the K18 residue and no modification (Pr) at the K23 residue. Detailed information is provided in the method section. **(B)** Mapping and quantification of UHRF1 ubiquitination target sites in H3 N-terminal tails and dependence of the ubiquitination on R2. GFP-H3 constructs carrying R2A, K18A, K23A or K18A-K23A mutations were transiently co-expressed in HEK 293T cells with HA-ubiquitin and after immunoprecipitation with the GFP-Trap, ubiquitinated GFP-H3 was detected by western blot with an anti-HA antibody. Equal loading of GFP-H3 is shown by the anti-GFP blot below. I, input; B, bound. Quantifications were performed with ImageJ. Shown are mean values \pm SEM of four to five independent experiments.

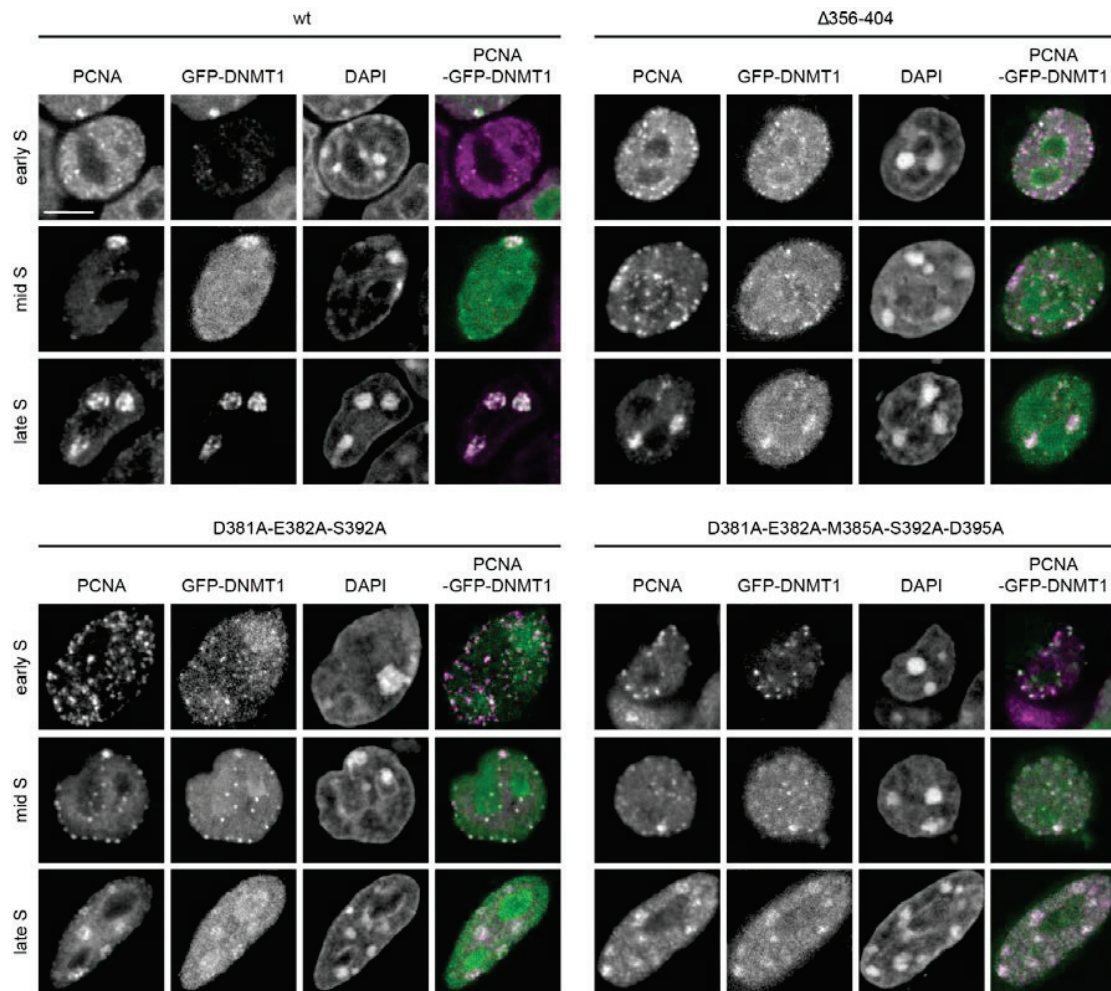


Supplementary information, Figure S6 DNMT1 UIM deletion and point mutants are defective in ubiquitinated H3 binding. (A)

Primary sequence alignment of the ubiquitin interacting motif (UIM) in the TS domain of DNMT1 from different species. The consensus sequence for single-sided UIMs [8] and the UIM mutations used in this study (D381A-E382A-S392A, D381A-E382A-M385A-S392A-D395A) are indicated. **(B)** Ubiquitinated histone H3 binding assay. After extraction of mononucleosomes from HEK 293T cells, the extracts were incubated with GFP-DNMT1 wt or mutants immobilized on the GFP-Trap and the bound fractions were analyzed with specific anti-H3, anti-ubiquitin and anti-GFP antibodies. **(C)** Characterization of the anti-H2AK119ub antibody (New England Biolabs) in comparison to the antibody against non ubiquitinated H2A (Millipore) by immunoblotting. 2.5 or 5 μg of acid extracted histones were loaded to analyze the specificity of the antibodies. H2Aub1, monoubiquitinated H2A; H2Aub2, diubiquitinated H2A. **(D)** Identification of the H2A 101-120 peptide carrying ubiquitination (GG) on K119 by LC-MS/MS. The sample was prepared by excision of a band from acid extracted histones corresponding to H2Aub1 in the immunoblot from (C). MS2 fragmentation spectrum of the precursor ion is shown in the inset. The displayed b and y ion series generated by CID fragmentation of the H2A 101-120 peptide modified with GG on K119 provided highly confident identification and post translational modification localization. Inset: mass, charge and measurement error determination of the H2A peptide of interest. Displayed is the isotopic distribution of the H2A peptide from which the mass-to-charge ratio (m/z) and the charge (2+) were obtained. Δm : difference between the expected and the measured masses, R : resolution of the MS measurement.



Supplementary information, Figure S7 DNMT1 UIM deletion and point mutants still interacting with UHRF1, affect maintenance DNA methylation. **(A)** Local DNA methylation analyses of *Dnmt1*^{-/-} ESCs transiently expressing GFP-DNMT1 wt or $\Delta 356-404$ and UIM point mutants (D381A-E382A-S392A, D381A-E382A-M385A-S392A-D395A). CpG methylation levels at the minor satellite repeats and the *Dnmt1o* promoter were analyzed by bisulfite treatment of genomic DNA, PCR amplification and direct pyrosequencing. Methylation levels of untransfected J1 *Dnmt1*^{-/-} cells are shown for comparison. Mean values \pm SD of two to four independent experiments were calculated. **(B)** Co-immunoprecipitation assay of GFP-TS and Ch-UHRF1 *in vitro*. GFP-DNMT1 TS domain wt or UIM mutants were co-expressed with Ch-UHRF1 and after co-immunoprecipitation using the GFP-Trap, the bound fractions were detected by western blot with specific antibodies against GFP and Ch. GFP was used as negative control. I, input; B, bound.



Supplementary information, Figure S8 DNMT1 UIM mutants show only weak association with PCNA-stained replication sites especially in late S phase. Confocal mid sections of GFP-DNMT1 wt or $\Delta 356-404$ and UIM point mutants (D381A-E382A-S392A, D381A-E382A-M385A-S392A-D395A) transiently expressed in J1 *Dnmt1*^{-/-} ESCs with respect to replicating DNA stained with a specific anti-PCNA antibody. Chromatin was counterstained with DAPI. Scale bar, 5 μ m.

Supplementary information, Table S1 PCR primers for bisulfite PCR and pyrosequencing.

Name	Sequence
F-major satellites	5' AAAATGAGAAATATTTATTTG 3'
R-major satellites biotinylated	5' GAGAAATATATACTTTAGGA 3'
F-outer <i>skeletal α-actin</i>	5' AGTTGGGGATATTTTTTATA 3'
R-outer <i>skeletal α-actin</i>	5' TGGGAAGGGTAGTAATATTT 3'
F-inner <i>skeletal α-actin</i>	5' TTTTGGTTAGTGTAGGAGAT 3'
R-inner <i>skeletal α-actin</i> biotinylated	5' TGGGAAGGGTAGTAATATTT 3'
F- <i>Dnmt1o-2-1</i>	5' GTTGTTTTTTGGTTTTTGTGGGTAT 3'
R-outer <i>Dnmt1o-2-2</i>	5' CAACCTTAACAACACAACATAAATA 3'
R-inner <i>Dnmt1o-2-3</i> biotinylated	5' CAACTATACACTATCAAATAACCTA 3'
F-minor satellites	5' TAATGAGTTATAATGAGAAA 3'
R-minor satellites biotinylated	5' ATATACACTATTCTACAAAT 3'

Supplementary information, Table S2 Sequence of histone H3 1-20 peptides used for the histone peptide pull-down and histone peptide binding assays. X3, trimethylated lysine; Z, acetylated lysine; X2a, asymmetrically dimethylated arginine.

Name	Sequence	Label
H3	ARTKQTARKSTGGKAPRKQLK	Biotin at C-terminus
H3K9me3	ARTKQTARX3STGGKAPRKQLK	
H3K9ac	ARTKQTARZSTGGKAPRKQLK	
H3	ARTKQTARKSTGGKAPRKQLK	TAMRA at C-terminus
H3K9me3	ARTKQTARX3STGGKAPRKQLK	
H3R2me2	AX2aTKQTARKSTGGKAPRKQLK	

Supplementary information, Table S3 DNA oligonucleotides used for the preparation of double stranded probes for *in vitro* DNA binding assays with the primer extension reaction. M, 5-methylcytosine.

Oligo name	DNA sequence
CG-up	5' CTCAACAACAACTACCATCCGGACCAGAAGAGTCATCATGG 3'
MG-up	5' CTCAACAACAACTACCATCMGGACCAGAAGAGTCATCATGG 3'
Fill-In-550	5' ATTO550-CCATGATGACTCTTCTGGTC 3'
Fill-In-647N	5' ATTO647N-CCATGATGACTCTTCTGGTC 3'
Substrate	Combination of oligos in the primer extension reaction
UMB-550	CG-up + Fill-In-550
HMB-647N	MG-up + Fill-In-647N

Results

2.3 A modular open platform for systematic functional studies under physiological conditions

A modular open platform for systematic functional studies under physiological conditions

Christopher B. Mulholland¹, Martha Smets¹, Elisabeth Schmidtmann¹, Susanne Leidescher¹, Yolanda Markaki¹, Mario Hofweber¹, Weihua Qin¹, Massimiliano Manzo¹, Elisabeth Kremmer², Katharina Thanisch¹, Christina Bauer¹, Pascaline Rombaut³, Franz Herzog³, Heinrich Leonhardt^{1,*} and Sebastian Bultmann^{1,*}

¹Ludwig Maximilians University Munich, Department of Biology II and Center for Integrated Protein Science Munich (CIPSM), Großhaderner Strasse 2, 82152 Planegg-Martinsried, Germany, ²Helmholtz Center Munich, German Research Center for Environmental Health (GmbH), Institute of Molecular Immunology, Marchioninistrasse 25, 81377 Munich, Germany and ³Gene Center and Department of Biochemistry, Ludwig Maximilians University Munich, Feodor-Lynen-Strasse 25, 81377 Munich, Germany

Received April 13, 2015; Revised May 13, 2015; Accepted May 14, 2015

ABSTRACT

Any profound comprehension of gene function requires detailed information about the subcellular localization, molecular interactions and spatio-temporal dynamics of gene products. We developed a multifunctional integrase (MIN) tag for rapid and versatile genome engineering that serves not only as a genetic entry site for the Bxb1 integrase but also as a novel epitope tag for standardized detection and precipitation. For the systematic study of epigenetic factors, including *Dnmt1*, *Dnmt3a*, *Dnmt3b*, *Tet1*, *Tet2*, *Tet3* and *Uhrf1*, we generated MIN-tagged embryonic stem cell lines and created a toolbox of prefabricated modules that can be integrated via Bxb1-mediated recombination. We used these functional modules to study protein interactions and their spatio-temporal dynamics as well as gene expression and specific mutations during cellular differentiation and in response to external stimuli. Our genome engineering strategy provides a versatile open platform for efficient generation of multiple isogenic cell lines to study gene function under physiological conditions.

INTRODUCTION

In the last decades targeted gene disruption has been a widely used approach to gain first insights into gene function. However, gene disruption studies are often hampered by high functional redundancy in mammalian systems and yield little information about the subcellular localization,

interactions and spatio-temporal dynamics of gene products. In order to gain comprehensive understanding of gene function these studies need to be complemented by more complex genetic manipulations such as fluorophore knockin, specific domain deletions or introduction of point mutations. Additionally, a systematic analysis of gene function requires application of biochemical as well as imaging techniques, which usually rely on the generation of gene specific antibodies, a technically demanding and time-consuming process.

Recently, RNA guided endonucleases (RGENs) derived from the prokaryotic Type II CRISPR/Cas (clustered regularly interspaced short palindromic repeats/CRISPR-associated) system have emerged as promising tools for the manipulation and modification of genetic sequences (1–4).

The specificity of RGENs is mediated by small guide RNAs (gRNAs) that bind to 20 bp within the target sequence and recruit the Cas9 nuclease to introduce a double strand break. Although this two-component system has greatly facilitated the generation of gene disruptions in bacteria, plants and mammals, concerns have been raised about considerable off-target effects (5–7). Furthermore, the low frequency of homologous recombination in mammals makes insertion of exogenous components such as fluorophore tags difficult and time-consuming.

In addition to RGENs, phage-derived serine integrases have received considerable attention as novel tools for genome engineering. Recently, Bxb1 was shown to have the highest accuracy and efficiency in a screen of fifteen candidate serine integrases tested in mammalian cells (8). Serine integrases are unidirectional, site-specific recombinases that promote the conservative recombination between phage attachment sites (*attP*) and bacterial attachment sites (*attB*)

*To whom correspondence should be addressed. Tel: +49 89 2180 74233; Fax: +49 89 2180 74236; Email: bultmann@bio.lmu.de
Correspondence may also be addressed to Heinrich Leonhardt. Tel: +49 89 2180 74232; Fax: +49 89 2180 74236; Email: h.leonhardt@lmu.de;

(9) with much higher recombination efficiencies (up to 80%) than the commonly used bidirectional tyrosine integrases, Cre or Flp (9–12).

In this study, we aim to combine the advantages of both RGES and unidirectional integrases into one fast, widely applicable and flexible method. We developed a novel strategy for genome engineering based on a CRISPR/Cas assisted in-frame insertion of an *attP* site, which we refer to as the multifunctional integrase (MIN) tag. At the genetic level, the MIN-tag serves as an attachment site for the serine integrase Bxb1 that can be used to introduce a broad range of prefabricated functional cassettes into the genomic locus with high specificity and efficiency. At the protein level, the MIN-tag functions as a novel epitope tag that can be detected with a highly specific monoclonal antibody and used for immunoprecipitation as well as immunofluorescence experiments. To demonstrate the versatility of the strategy, we generated MIN-tagged murine embryonic stem cell (mESC) lines for a variety of major epigenetic factors, including *Dnmt1*, *Dnmt3a*, *Dnmt3b*, *Tet1*, *Tet2*, *Tet3* and *Uhrf1*. We created a toolbox of vectors for Bxb1-mediated recombination to generate isogenic cell lines harboring knockout cassettes, fluorescent protein fusions, enzymatic tags and specific mutations; all derived from a single entry cell line ensuring maximal biological comparability. We demonstrate the power of this strategy using proximity-dependent protein labeling to identify novel interactors of TET1 in mESCs as well as to systematically study the subcellular localization, binding kinetics and protein expression dynamics of the *de novo* methyltransferase DNMT3B during epiblast differentiation.

MATERIALS AND METHODS

Western blotting and immunoprecipitation

Western blot analysis was performed using the following primary antibodies: anti-DNMT1, anti-DNMT3a (Imgenex, 64B1446); anti-DNMT3b (Abcam, 52A1018); anti-UHRF1 (13); anti-TET1, anti-TET2 and anti-TET3 (14); anti-GFP antibody (Roche, 11814460001); anti- β -Actin (Sigma, A5441); anti-SNF2H (Abcam, ab22012). Blots were probed with anti-rat (Jackson ImmunoResearch, 112-035-068), anti-mouse (Sigma, A9044) and anti-rabbit (Biorad, 170–6515) secondary antibodies conjugated to horseradish peroxidase (HRP) and visualized using an ECL detection kit (Pierce). An anti-mouse antibody conjugated to Alexa 488 (Life Technologies, A21202) was used for fluorescence detection of western blots using the Typhoon 9400 (GE Healthcare) imaging system.

For immunoprecipitation, $\sim 1 \times 10^6$ *Dnmt1^{attP/attP}*, *Dnmt3b^{attP/attP}* or wt cells were harvested in ice cold phosphate buffered saline (PBS), washed twice and subsequently homogenized in 200 μ l lysis buffer (20 mM Tris/HCl pH 7.5, 150 mM NaCl, 0.5 mM EDTA, 1 mM PMSF, 0.5% NP40). After centrifugation (10 min, 14 000 g, 4°C) the supernatant was adjusted with dilution buffer (20 mM Tris/HCl pH 7.5, 150 mM NaCl, 0.5 mM EDTA, 1 mM PMSF) to a final volume of 300 μ l. A total of 50 μ l were mixed with sodium dodecyl sulphate (SDS)-containing sample buffer (referred to as input (I)). For pull-downs, 100 μ l (4 μ g) of either 5A10 DNMT1 antibody (15) or

the newly generated MIN-tag antibody 1E1 was added to the cell lysates and incubated 2 h at 4°C. For pull-down of immunocomplexes, 40 μ l of protein G agarose beads (GE Healthcare, Freiburg, Germany) equilibrated in dilution buffer were added and incubation continued for 2 h. After centrifugation (2 min, 5000 \times g, 4°C) 50 μ l of the supernatant was collected (referred to as flow-through (FT)) while the remaining supernatant was removed. The beads were washed twice with 1 ml dilution buffer containing 300 mM NaCl. After the last washing step, the beads were resuspended in 50 μ l Laemmli buffer and boiled for 10 min at 95°C. For immunoblot analysis, 3% of the input and the flow-through as well as 30% of the bound (B) fraction were separated on a 10% sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) and subjected to western blot analysis.

Immunofluorescence staining and microscopy

Immunostaining was performed as described previously (16). Briefly, cells cultured on coverslips were fixed with 4% paraformaldehyde for 10 min, washed with PBST (PBS, 0.02% Tween20) and permeabilized with PBS supplemented with 0.5% Triton X-100. Both primary and secondary antibody were diluted in blocking solution (PBST, 2% BSA, 0.5% fish skin gelatin). Coverslips with cells were incubated with primary and secondary antibody solutions in dark humid chambers for 1 h at RT; washings after primary and secondary antibodies were done with PBST. Following secondary antibody incubations, cells were post-fixed with 4% paraformaldehyde for 10 min. For DNA counterstaining, coverslips were incubated in a solution of DAPI (2 μ g/ml) in PBS. Coverslips were mounted in antifade medium (Vectashield, Vector Laboratories) and sealed with colorless nail polish.

For immunolabeling, the following primary antibodies were used: anti-DNMT1 (15); anti-DNMT3A (Imgenex, 64B1446); anti-DNMT3B (Abcam, 52A1018); anti-UHRF1 (13); anti-TET1, anti-TET2 (14); GFP-Booster.ATTO488 (Chromotek). The secondary antibodies were anti-rabbit conjugated to DyLight fluorophore 594 (Jackson ImmunoResearch, 711-505-152), anti-mouse conjugated to Alexa 488 (Life Technologies, A21202), anti-rat conjugated to Alexa 488 (Life Technologies, A21208) or Alexa 594 (Life Technologies, A21209).

Single optical sections or stacks of optical sections were collected using a Leica TCS SP5 confocal microscope equipped with Plan Apo 63 \times /1.4 NA oil immersion objective and lasers with excitation lines 405, 488, 561 and 633 nm.

Live cell imaging experiments were performed on an UltraVIEW VoX spinning disc microscope assembled to an Axio Observer D1 inverted stand (Zeiss) and using a 63 \times /1.4 NA Plan-Apochromat oil immersion objective. The microscope was equipped with a heated environmental chamber set to 37°C and 5% CO₂. Fluorophores were excited with 488 nm or 561 nm solid-state diode laser lines. Confocal image series were typically recorded with 14-bit image depth, a frame size of 1024 \times 1024 pixels and a pixel size of 110 nm. z-stacks of 12 μ m with a step size of 1 μ m were recorded every 30 min for about 24 h or for the live

cell series of *Dnmt3b^{attP/attP}* every hour for 60 h. To avoid photodamage of the cells, the AOTF of the laser was set to low transmission values of 6–10%. Binning was set to 2×.

Super-resolution microscopy

Super-resolution images were obtained with a DeltaVision OMX V3 3D-SIM microscope (Applied Precision Imaging, GE Healthcare), equipped with a 60×/1.42 NA PlanApo oil objective and sCMOS cameras (Olympus). A z-step size of 125 nm was used during acquisition. SI raw data were reconstructed and deconvolved with the SoftWorX 4.0 software package (Applied Precision). FIJI and Photoshop CS5.1 (Adobe) were used for image processing and assembly.

Antigen preparation, immunization, generation of hybridomas and ELISA screening

For the translated attP peptide, the MIN antigen (attP peptide) was designed with the following sequence SGQPPRSQWCTVQT-Cys. Peptides were synthesized, HPLC purified and coupled to OVA (Peps4LifeSciences-Anette Jacob; Heidelberg). Lou/c rats were immunized subcutaneously and intraperitoneally with a mixture of 50 µg peptide-OVA, 5 nmol CPG oligonucleotide (Tib Molbiol, Berlin), 500 µl PBS and 500 µl incomplete Freund's adjuvant. A boost without adjuvant was given 6 weeks after primary injection. Fusion of the myeloma cell line P3 × 63-Ag8.653 with the rat immune spleen cells was performed using polyethylene glycol 1500 (PEG 1500, Roche, Mannheim, Germany). After fusion, the cells were plated in 96 well plates using RPMI1640 with 20% fetal calf serum, penicillin/streptomycin, pyruvate, non-essential amino acids (Gibco) supplemented by hypoxanthine-aminopterin-thymidine, (HAT) (Sigma, St Louis, MO, USA). Hybridoma supernatants were tested in a solid-phase immunoassay. Microliter plates were coated with avidin (3 µg/ml, Sigma) over night. After blocking with 2% FCS in PBS, plates were incubated with biotinylated MIN peptide at a concentration of 0.2 µg/ml in blocking buffer. After washing the plates, the hybridoma supernatants were added. Bound rat mAbs were detected with a cocktail of HRP-labeled mouse mAbs against the rat IgG heavy chains, thus avoiding IgM mAbs (α-IgG1, α-IgG2a, α-IgG2b (ATCC, Manassas, VA, USA), α-IgG2c (Ascension, Munich, Germany). HRP substrate conversion was visualized with ready to use TMB (1-Step™ Ultra TMB-ELISA, Thermo). MIN-tag clone 1E1 (rat IgG1) was stably subcloned and further characterized.

A set of 25 rat derived hybridoma supernatants were tested for specificity against an integrated attP peptide in the *Dnmt1* locus using both western blot analysis and high content microscopy. Western blots were prepared as mentioned previously. Each supernatant was used in a 1:10 dilution. Blots were probed with an anti-rat secondary antibody conjugated to HRP.

Cells were prepared for immunofluorescence as described above, with the exception that cells were fixed on a 96-well Cell Carrier® plate (Greiner). Cells in individual wells were incubated with the various hybridoma supernatants (1:100)

for 1 h. As a secondary antibody, anti-rat conjugated to Alexa 488 (Life Technologies, A21208) was used. Nuclei were counterstained using DAPI. Images of stained cells were acquired automatically with an Operetta high-content imaging system using a 40× air objective (PerkinElmer). DAPI and ATTO488 coupled antibodies were excited and their emissions recorded using standard filter sets. Exposure times were 10 and 400 ms for DAPI and ATTO488, respectively. All monoclonal antibodies described in this study are available upon request.

The MIN antibody are available via http://human.bio.lmu.de/_webtools/MINtool/AB_info.html.

DNA methylation analysis

For the analysis of DNA methylation levels, genomic DNA was isolated using the QIAamp DNA Mini Kit (QIAGEN). Bisulfite treatment was performed using the EZ DNA Methylation-Gold™ Kit (Zymo Research Corporation) according to the manufacturer's protocol. Subsequently, the major satellite repeats sequence was amplified using the primers described in (17). The biotinylated polymerase chain reaction (PCR) products of the second PCR were analyzed by pyrosequencing (Varionostic GmbH, Ulm, Germany).

Targeting donor and plasmid construction

Plasmid sequences can be found in Supplementary Table S6. Targeting donor constructs were either synthesized as ssDNA oligonucleotides (Integrated DNA Technologies) or produced by amplifying 300 to 200 bp long homology arms with the respective external and internal primer sets (Supplementary Table S2). These PCR products of the 5' and 3' homology arms were pooled and an overlap extension PCR with the external primers was performed to yield the final targeting fragments. The gRNA vector was synthesized at Eurofins MWG Operon based on the sequences described (3). The subcloning of targeting sequences was performed by circular amplification. The surrogate reporter (pSR) was generated by inserting *in vitro* annealed DNA oligos via *Asi*SI and *Nru*I into pCAG-mCh (18). eGFP was amplified using the primers eGFP-F and eGFP-R and sequentially cloned into pCAG-mCh-*Nru*I linker to generate the pSR construct. Reporters were generated by subcloning *in vitro* annealed DNA oligos containing CRISPR target sites into *Kpn*I and *Nhe*I digested pSR. The attB-GFP-knockin construct was generated from R6K-NFLAP (19) by ligation free cloning (20) rearranging the backbone sequences into the artificial intron and introducing the attB site 5' of the GFP open reading frame (ORF), removing its start codon. The attB-GFP-Poly(A) and attB-mCh-Poly(A) constructs were created by amplifying the GFP ORF including the stop codon and SV40 Poly(A) signal from pCAG-eGFP-IB and inserted into the attB-LAP-tag backbone by ligation free cloning. The attB-mCh-Poly(A)-mPGK-PuroR construct was generated by subcloning the mPGK-PuroR sequence from pPthc-Oct3/4 (21) and ligating it into the *Eco*RV site of the attB-mCh-Poly(A) construct. The attB-GFP-Poly(A)-mPGK-NeoR was produced by first exchanging

the PuroR in pPthc-Oct3/4 with NeoR from pEGFP-C1 (22) using HindIII. The combined mPGK-NeoR was then subcloned into the attB-GFP-Poly(A) vector via the same EcoRV site mentioned previously. The attB-GFP-Dnmt1cDNA-Poly(A), attB-GFP-Tet1cDNA-Poly(A) and attB-GFP-Dnmt3b1cDNA-Poly(A) constructs were generated by inserting the appropriate cDNAs from constructs reported previously (17,23–24) via AsiSI/NotI sites into the attB-GFP-Poly(A) and attB-mCh-Poly(A) vectors respectively. The attB-GFP-Dnmt3b6-Poly(A), attB-GFP-Tet1-d1–1363-Poly(A), attB-GFP-Tet1-d833–1053-Poly(A), attB-GFP-Tet1-d833–1363-Poly(A) vectors were produced via circular amplification with overlap extension primers using the above mentioned attB-GFP-Dnmt1/Dnmt3b1/Tet1cDNA-Poly(A) constructs as templates.

The attB-GFP-Dnmt3b6-Poly(A)-mPGK-NeoR and attB-mCh-Dnmt3b1-Poly(A)-mPGK-PuroR integration constructs were created by inserting the Dnmt3b6 and Dnmt3b1 sequences (from attB-GFP-Dnmt3b6-Poly(A) and attB-GFP-Dnmt3b1-Poly(A)) using AsiSI/NotI sites into attB-GFP-Poly(A)-mPGK-NeoR and attB-mCh-Poly(A)-mPGK-PuroR vectors, respectively.

All constructs described in this study are available via Addgene or via http://human.bio.lmu.de/_webtools/MINTool/.

Cell culture

J1 ESCs were maintained on gelatin-coated dishes in Dulbecco's modified Eagle's medium supplemented with 16% fetal bovine serum (FBS, Biochrom), 0.1 mM β -mercaptoethanol (Invitrogen), 2 mM L-glutamine, 1 \times MEM Non-essential amino acids, 100 U/ml penicillin, 100 μ g/ml streptomycin (PAA Laboratories GmbH), 1000 U/ml recombinant mouse LIF (Millipore) and 2i (1 μ M PD032591 and 3 μ M CHIR99021 (Axon Medchem, Netherlands), referred to as ESC medium. Differentiation of naive pluripotent stem cells to epiblast-like cells was performed according to the protocol of (25). Briefly, J1 ESCs were maintained in the ground state in Geltrex (Life Technologies) coated flasks and cultured in N2B27 (50% neurobasal medium (Life Technologies), 50% DMEM/F12 (Life Technologies), 2 mM L-glutamine (Life Technologies), 0.1 mM β -mercaptoethanol, N2 supplement (Life Technologies), B27 serum-free supplement (Life Technologies) containing 2i and 1000 U/ml LIF 100 U/ml Penicillin-streptomycin) for at least three passages before differentiation. To differentiate naive ESCs into epiblast-like cells, cells were replated in N2B27 differentiation medium containing 10 ng/ml Fgf2 (R&D), 20 ng/ml Activin A (R6D) and 0.1 \times Knockout Serum Replacement (KSR)(Life Technologies). Time point 0 h in differentiation time-course experiments corresponds to the time N2B27 differentiation medium was added to cells.

Generation of MIN-tagged and Bxb1-mediated knockin cell lines

To produce MIN-tagged cell lines, 5×10^5 cells were dissociated and seeded in 0.2% gelatin (Sigma-Aldrich) coated p35 plates. After 3 h, cells were transfected with 2 μ g of

the MIN-tag donor/homology ssDNA oligo or PCR product, 0.5 μ g gRNA construct, 0.5 μ g surrogate reporter construct and 1 μ g Cas9 using Lipofectamine 3000 (Invitrogen) according to the manufacturer's instructions. For Bxb1-mediated recombination of attB constructs, 5×10^5 cells were transfected with 1 μ g pCAG-NLS-HA-Bxb1 expression plasmid ((26) addgene 51271), 1 μ g of the respective attB construct and 0.5 μ g Bxb1 surrogate reporter. For both MIN-Tagging and Bxb1-mediated recombination, cells were dissociated, resuspended in ESC medium 48 h post transfection and then analyzed and sorted with a FACS Aria II (Becton Dickinson). For MIN-tagging, enrichment of cells with RGEN activity was accomplished by single-cell sorting GFP and mCherry positive cells into 96-well plates (Falcon) containing 150 μ l of ESC medium. For Bxb1-mediated recombination, cells with Bxb1 activity were enriched for by single-cell sorting GFP positive cells into 96-well plates. Alternatively for Bxb1-mediated integration using antibiotic selection, cells were replated into p150 plates with ESC medium containing G418 (0.5 mg/ml, AppliChem) and puromycin (1 μ g/ml, AppliChem) 48 h post transfection.

Identification of MIN-tagged and Bxb1-mediated knockin cell lines with restriction fragment analysis and PCR screening

After ~7 days (until colonies were readily visible), plates from single-cell sortings were screened for colony growth. Surviving colonies were dissociated and individually replated onto two 96-well plates. Genomic DNA was isolated from one plate after 2–3 days, while the second plate remained in culture. To identify MIN-tagged clones, the region surrounding the ATG (or stop codon in the case of C-terminal tagging) was PCR amplified using the appropriate external and screening primers (Supplementary Table S2). For restriction fragment analysis, 10 μ l of these PCR products were digested with either HincII or SacII and then analyzed on 1.5% agarose gels. PCRs of positive clones were confirmed by Sanger sequencing. To screen for Bxb1-mediated recombination, we employed a three-primer PCR strategy using the respective external primers flanking the MIN-tagged locus and an attL-specific primer (Supplementary Figure S3A, Table S2). For Bxb1-mediated integrations using antibiotic selection, mESC colonies were picked, dissociated using trypsin and plated into individual wells on 96-well plates ~7 days after starting antibiotic selection. Genomic DNA isolation and screening PCRs were performed as described above. Clones harboring the desired MIN-tag insertion or Bxb1-mediated integration were expanded, frozen and stored in liquid nitrogen.

All cell lines are available at http://human.bio.lmu.de/_webtools/MINTool/cell_lines.html.

Genomic DNA isolation for PCR

Cells were lysed in multi-well plates by the addition of 50 μ l lysis buffer (10mM Tris/HCl pH 7.4, 10mM EDTA, 10mM NaCl, 50 μ g/ml Proteinase K, 1.7 μ M SDS) per well. The Plates were subsequently incubated at -80°C for 15 min, followed by 3 h at 56°C . Heat inactivation of Proteinase K

was performed by incubation at 85°C for 20 min. The resulting crude DNA lysates were directly subjected to PCR.

BioID

BioID experiments were performed after (27) using extracted crude nuclei (adapted from (28)) as input material. In brief, cells were cultured for 48 h with or without addition of 50 μ M biotin. Cell pellets ($\sim 4 \times 10^7$ cells) were washed once in buffer A (10 mM HEPES/KOH pH 7.9, 10 mM KCl, 1.5 mM MgCl₂) and resuspended in buffer A containing 0.15% NP-40 and 1 \times protease inhibitor (SERVA). Samples were homogenized using a pellet pestle. After centrifugation, crude nuclei pellets were washed once with PBS. Crude nuclei were resuspended in BioID-lysis buffer (0.2% SDS, 50 mM Tris/HCl pH 7.4, 500 mM NaCl, 1 mM DTT, 1 \times protease inhibitor), supplemented with 2% Triton X-100 and subjected to sonication twice using a Branson Sonifier 450 (15% amplitude, 0.3 s pulse, 0.6 s pause, total time 30 s). Samples were diluted 1:1 with 50 mM Tris/HCl pH 7.4 after the first sonication step. Pulldown of biotinylated proteins was performed overnight at 4°C with rotation using M-280 Streptavidin Dynabeads (Life Technologies) for subsequent mass spectrometry or Streptactin-Superflow agarose beads (IBA) for SDS-PAGE analysis, respectively. Beads were washed with wash buffer 1 (2% SDS), wash buffer 2 (0.1% desoxycholic acid, 1% Triton X-100, 1 mM EDTA, 500 mM NaCl, 50 mM HEPES/KOH pH 7.5) and wash buffer 3 (0.5% desoxycholic acid, 0.5% NP-40, 1 mM EDTA, 500 mM NaCl, 10 mM Tris/HCl pH 7.4) followed by two washing steps with 50 mM Tris/HCl pH 7.4. For SDS-PAGE analysis, proteins were silverstained after (29).

Digest of proteins and sample preparation for LC-MS/MS

On-beads digest of proteins was performed as described in (28). All steps were carried out at room temperature. Beads were resuspended in 2 M Urea in Tris/HCl pH 7.5, reduced with 10 mM DTT for 20 min and subsequently alkylated with 50 mM chloroacetamide for 20 min. A total of 0.25 μ g Pierce Trypsin Protease (Thermo Scientific) was added for 2 h. Beads were collected by centrifugation and the resulting peptide supernatant was further incubated overnight with addition of 0.1 μ g trypsin. Peptides were desalted using StageTips (30).

LC-MS/MS and data analysis

Peptides were reconstituted in 20 μ l mobile phase A (2% v/v acetonitrile, 0.1% v/v formic acid) and analyzed by tandem mass spectrometry using a EASY-nLC 1000 nano-HPLC system connected to a LTQ Orbitrap Elite mass spectrometer (Thermo Fisher Scientific). About 2–4 μ l of the peptide mixture were separated onto a PepMap RSLC column (75 μ m ID, 150 mm length, C18 stationary phase with 2 μ m particle size and 100 Å pore size, Thermo Fisher Scientific) and introduced into the mass spectrometer at a flow rate of 300 nl/min running a gradient from 5 to 35% mobile phase B (98% v/v acetonitrile, 0.1% v/v formic acid). Ion source and transmission parameters of the mass spectrometer were set to spray voltage = 2 kV, capillary temperature = 275°C. The mass spectrometer was operated in

data-dependent mode, selecting up to 10 precursors from a MS1 scan (resolution = 60 000) in the range of m/z 250–1800 for collision-induced dissociation (CID). Singly (+1) charged precursor ions and precursors of unknown charge states were rejected. CID was performed for 10 ms using 35% normalized collision energy and the activation q of 0.25. Dynamic exclusion was activated with a repeat count of one, exclusion duration of 30 s, list size of 500 and the mass window of ± 10 ppm. Ion target values were 1 000 000 (or maximum 10 ms fill time) for full scans and 10 000 (or maximum 100 ms fill time) for MS/MS scans, respectively. Raw data were analyzed using MaxQuant Version 1.5.2.8 (31) using the MaxLFQ label free quantification algorithm (32) and the match-between-runs functionality. UniprotKB MOUSE.fasta was used as a reference database (33). A maximum of two missed cleavages and a false discovery rate of 1% were set as parameters. Oxidation of methionine and biotinylation were searched as variable modifications and carbamidomethylation of cysteine residues as fixed modification. For statistical analysis, the Perseus software version 1.5.1.6 was used (31). Significance was tested using a two sided Student's *t*-test and a permutation based FDR calculation. GO enrichment analysis was performed with the *Gene Ontology enrichment and Lysis and visualization tool* (GORilla, (34)). A *P*-value < 0.01 was considered significant.

FRAP

Live cell imaging and FRAP experiments were typically performed on an UltraVIEW VoX spinning disc microscope with integrated FRAP PhotoKinesis accessory (PerkinElmer) assembled to an Axio Observer D1 inverted stand (Zeiss) and using a 63 \times /1.4 NA Plan-Apochromat oil immersion objective. The microscope was equipped with a heated environmental chamber set to 37°C. Fluorophores were excited with 488 nm (exposure time: 400 ms, laser power: 15%) or 561 nm (exposure time: 450 ms, laser power: 30%) solid-state diode laser lines. Confocal image series were typically recorded with 14-bit image depth, a frame size of 256 \times 256 pixels and a pixel size of 110 nm. For photobleaching experiments, the bleach regions, typically with a diameter of 2 μ m, were manually chosen to cover the chromocenters. Photobleaching was performed using one iteration with the acousto-optical tunable filter (AOTF) of the 488 nm laser line set to 100% transmission. Typically, 10 pre-bleach images were acquired at a rate of 1 s per timepoint and 60 post-bleach frames were recorded at a rate of 10 s per timepoint. Data correction, normalization and quantitative evaluations were performed by automated processing with ImageJ (<http://rsb.info.nih.gov/ij/>) using a set of newly developed macros followed by calculations in Excel.

RESULTS

A fast and efficient strategy to generate MIN-tagged genomic loci

Our novel genome engineering strategy relies on the CRISPR/Cas-assisted insertion of the MIN-tag sequence into the open reading frame of a target gene either directly

downstream of the start codon or upstream of the stop codon (Figure 1A and Supplementary Figure S2H). Neither regulatory regions nor gene structure are altered, leading to preservation of the endogenous expression pattern and post-transcriptional processing of the gene of interest.

Since epigenetic processes undergo dramatic changes during early embryonic development and are tightly regulated, we tested the efficacy and versatility of our method by targeting the DNA modifying enzymes *Dnmt1*, *Dnmt3a*, *Dnmt3b*, *Tet1*, *Tet2* and *Tet3* as well as the chromatin binding protein *Uhrf1* in mESCs (Figure 1D). We generated targeting donors containing the 48 bp MIN-tag sequence flanked by short homology arms (200–300 bp for PCR-based donors or 76 bp for single stranded DNA oligos). We next designed specific gRNAs to target sequences located either in close proximity to or overlapping the start or stop codon of the respective genes. As scarless integration of the MIN-tag requires a resistance free selection strategy we used a surrogate reporter assay to enrich for cells that express an active Cas9:gRNA complex by fluorescence-activated cell sorting (FACS) (Figure 1B and C). In this reporter assay, the target sequence is inserted between the ORF of mCherry (mCh) and GFP thereby disrupting the reading frame of the fusion. GFP is expressed only when the target sequence is cleaved by a specific and active Cas9:gRNA complex, which causes small, frameshifting insertions or deletions by non-homologous end joining (NHEJ) restoring the reading frame of the fluorescent protein (35). For each targeting, we co-transfected mESCs with a mixture of surrogate reporter construct, gRNA vector, Cas9 expression plasmid and the specific targeting MIN-tag donor fragment. After single cell sorting of GFP positive cells and expansion of the resulting colonies, we isolated genomic DNA by a fast and simplified in-well lysis protocol to screen for positive clones by PCR and analytical restriction digest. This allows the identification of hetero- and homozygous insertions already at this stage (Supplementary Figure S1D). Combined, all targeting yielded positive clones with an average efficiency of 3% for homozygous and 1% for heterozygous insertions (Supplementary Table S1). All targeted genes were expressed normally and subcellular localization as well as enzymatic activity was not disrupted in comparison to wild-type (wt) cells (Supplementary Figures S1 and S2). In addition, the possibility of C-terminal tagging (see *Uhrf1* (C); Figure 1D and Supplementary Figure S2H) allows the MIN-tag to be used in cases where N-terminal targeting disturbs protein function.

Taken together, these results demonstrate that the MIN-tag can efficiently be integrated at precise genomic locations using a CRISPR/Cas assisted, fluorescence based selection strategy.

Generation of a highly specific monoclonal antibody recognizing the MIN epitope

Insertion of the MIN-tag into the ORF of target genes leads to expression of a small peptide that does not occur in the mammalian proteome (Figure 2A). This unique feature allowed us to generate a highly specific monoclonal antibody against MIN-tagged proteins. Immunofluorescence (IF) stainings of a mixed *Dnmt1^{attP/attP}* and wt culture dis-

tinguished single MIN-tagged cells and colonies from wt cells, demonstrating the high specificity of the anti-MIN antibody (Figure 2B). Pull-down experiments in *Dnmt1^{attP/attP}* cell extracts showed a quantitative enrichment of DNMT1 in the bound fraction (Figure 2C). Furthermore, pull-down of DNMT3B using the anti-MIN antibody efficiently coprecipitated SNF2H, a known interactor of DNMT3B, in protein extracts of *Dnmt3b^{attP/attP}* cells, but not in wt control extracts (Figure 2D) (36).

Collectively, these data show that the MIN-tag can be utilized as a universal epitope tag for IF and immunoprecipitation (IP), thus allowing the investigation of localization and molecular interactions of MIN-tagged proteins.

Functionalization of MIN-tagged genes by Bxb1-mediated recombination

To demonstrate the versatility of the MIN-tag as a Bxb1 integration site, we constructed a toolbox of functional cassettes, which we recombined into the MIN-tagged locus of the maintenance DNA methyltransferase *Dnmt1* (*Dnmt1^{attP/attP}*). First, we generated a knockout vector carrying the *attB* site directly in front of the ORF of GFP followed by a stop codon and a polyadenylation signal (*attB-GFP-Poly(A)*, Figure 3A) that we transfected together with a codon-optimized Bxb1 expression construct in the *Dnmt1^{attP/attP}* cell line. Successful recombination events were identified by GFP expression and single cells sorted by FACS (Figure 3B). We designed a multiplex PCR strategy that takes advantage of the unique *attL* site generated by successful recombination to facilitate identification of positive clones and their zygosity (Figure 3D and Supplementary Figure S3A). PCR screening of sorted clones revealed that the *attB-GFP-Poly(A)* construct had been successfully integrated into both alleles in 13 (56.5%) clones (Supplementary Table S3). Of those, we examined three clonal cell lines all of which exhibited no residual expression of DNMT1 by western blot analysis and IF (Figure 3F; Supplementary Figure S3B and C). For functional characterization, we analyzed DNA methylation levels at major satellite repeats, one of the main substrates for DNA methylation activity of DNMT1 during replication (37,38). Due to the loss of the maintenance DNA methyltransferase in the *Dnmt1^{KO/KO}* clones, a severe hypomethylation was observed at this sequence (Figure 3E). Taken together, our *attB-GFP-Poly(A)* vector proved to be a valuable tool to generate genetically-defined gene knockouts in MIN-tagged cell lines.

Second, we designed a GFP knockin construct that can be used to generate in-frame GFP fusions of MIN-tagged genes. To avoid disruption of the gene locus and preserve the endogenous splicing sites, we placed the bacterial backbone sequences into an artificial intron splitting the GFP ORF into two exons (19) (Figure 3A). After recombination and FACS sorting for GFP expressing cells, the GFP knockin construct integrated in both alleles of the *Dnmt1* locus in 13 clones (41.9%), without altering physiological DNMT1 expression levels (Figure 3G, Supplementary Figure S3D and Table S3). Live cell imaging of *Dnmt1^{GFP/GFP}* cells revealed a normal localization of GFP-DNMT1 throughout the cell cycle (15,24) (Supplementary Figure S3E), demonstrating

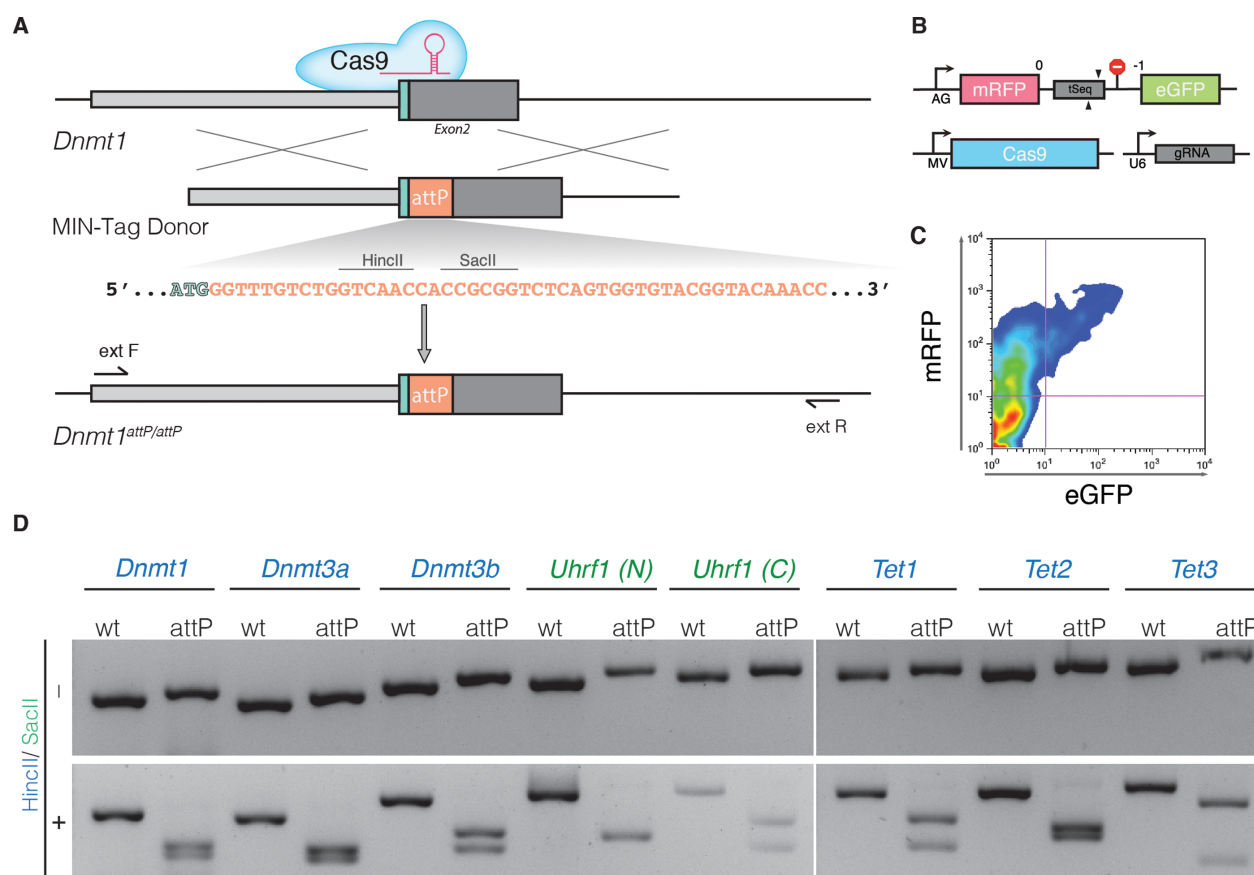


Figure 1. Generation of MIN-tagged cell lines. **(A)** Schematic overview of MIN-tag insertion into the *Dnmt1* locus via CRISPR/Cas assisted gene editing. The MIN-tag donor harbors the *attP* site and homology to the genomic sequence 5' and 3' of the start codon. Integration is facilitated by double strand breaks created by Cas9 directed to the target sequence by a specific gRNA. Restriction enzyme recognition sites used for screening in this study are indicated above the *attP* sequence. **(B)** Schematic overview of the surrogate reporter used to enrich for cells expressing a functional Cas9 complex. The respective Cas9 target sequence (tSeq) is placed downstream of mRFP followed by a stop codon and an out-of-frame GFP ORF. This surrogate reporter is transfected into the cells together with a vector expressing Cas9 and a U6 driven gRNA expression cassette. **(C)** Cells that express a functional Cas9 complex can then be identified by expression of GFP and enriched via FACS. **(D)** Screening PCRs followed by restriction digest with HincII or SacII of all generated MIN-tagged cell lines. (N) and (C) refer to N- and C-terminal tagging, respectively.

that DNMT1 regulation was not impaired. Albeit only at low frequencies, Bxb1 has been shown to damage recombination sites (8). Therefore, we sought to confirm that the Bxb1-mediated recombination of the GFP cassette at the MIN-tagged locus occurred without error via site-specific recombination. We sequenced the region flanking the *attL* site in the *Dnmt1*^{GFP/GFP} cell line (Supplementary Figure S4) and determined that the GFP cassette was accurately integrated in a scarless fashion. In summary, this attB-GFP vector is suited to express GFP fusion proteins from the endogenous promoter preserving physiological regulation and splicing of the target gene.

Finally, we investigated whether the MIN-tag can be used to generate cell lines expressing mutants of the target gene for functional screenings or disease modeling. We cloned the cDNA of *Dnmt1* into the *attB*-GFP-*Poly(A)* construct in-frame with GFP and performed recombination as described above. We identified 10 (66.6%) clones in which integration had occurred, of which 9 (60%) were homozygous for the *Dnmt1* cDNA knockin (Supplementary Table S3).

Expression analysis by western blot and live cell imaging revealed that the endogenous DNMT1 protein was completely replaced by the *Dnmt1* mini gene product and exhibited normal localization (Figure 3H, Supplementary Figure S3F).

All in all, we show that MIN-tagged entry cell lines can be efficiently functionalized with a flexible toolbox of attB-vectors to generate gene knockouts, N-terminal fusion constructs such as GFP and cDNA knockins. In total, we generated 15 derivatives of our MIN-tagged cell lines so far. The efficiency of Bxb1-mediated recombination ranged from 33 to 67%, with an average of 50% (Supplementary Table S3, Figure S5). This demonstrates the efficacy of our system as well as the simplicity with which MIN-tagged cell lines can be modified and functionalized by prefabricated cassettes. The error-prone step of CRISPR/Cas-mediated insertion of the MIN-tag is necessary only once to generate an entry cell line, which can then be specifically manipulated with a variety of recombination vectors, allowing maximum biological comparability.

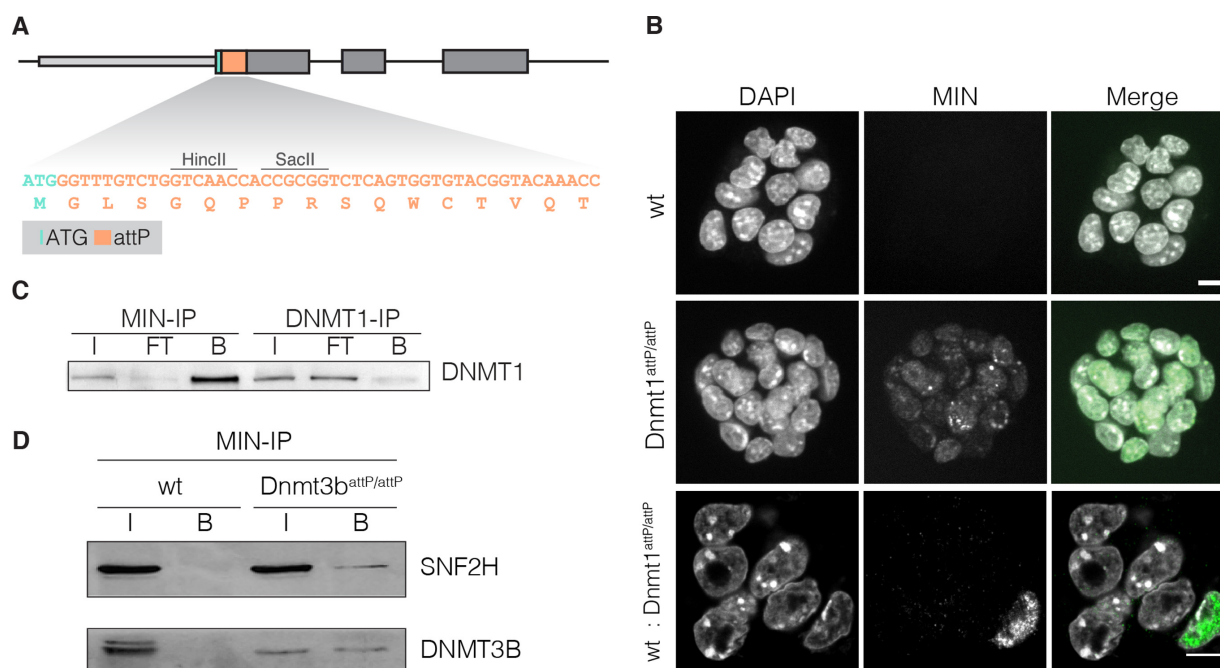


Figure 2. Application of the anti-MIN monoclonal antibody. (A) DNA sequence of the *attP* site and corresponding translated MIN peptide sequence (orange). (B) Fluorescence micrographs of wt mESCs, Dnmt1attP/attP cells and of a mixed culture (1:10) of wt and Dnmt1attP/attP cells stained with the anti-MIN antibody. DAPI is used as DNA counterstain. Scale bars represent 5 μ m. (C) IP experiments performed with anti-MIN and anti-DNMT1 antibody in Dnmt1attP/attP cell extracts (input (I), flow through (FT), bound (B)). (D) Co-IP of DNMT3B in wt and Dnmt3battP/attP cells using the anti-MIN antibody. DNMT3B co-precipitated SNF2H in Dnmt3battP/attP cells as determined by western blot.

Using the MIN-tag strategy to study endogenous protein regulation

As elucidating the function of uncharacterized protein domains requires systematic analysis, we generated a series of deletion constructs covering the N-terminus of TET1, which we aimed to recombine into our *Tet1^{attP/attP}* cell line (Figure 4A). However, we were unable to identify positive recombination events by FACS due to low expression of this target gene. To circumvent this problem, we developed a surrogate reporter system for Bxb1 mediated recombination that can be used to enrich for positive recombination events (Figure 3C). The Bxb1 surrogate reporter construct consists of a constitutive promoter followed by an *attP* site and a Poly(A) sequence. Upon transfection, Bxb1 mediates the recombination of a fluorophore (e.g. GFP) containing *attB* plasmid with the Bxb1 surrogate reporter, which results in the expression of GFP. This allows enrichment of positive recombination events, even when the MIN-tagged gene is not expressed or only at low levels.

Using the Bxb1 surrogate reporter for enrichment and the above described PCR strategy for screening, we were able to generate four *Tet1* knockin cell lines expressing N-terminal deletion constructs from the endogenous promoter. Western blot analysis revealed complete replacement of wt TET1 expression by the knockin constructs (Figure 4B). These cell lines can be used for future systematic studies of the regulatory function of the TET1 N-terminus that is largely unknown so far.

Taking advantage of the MIN-tag strategy to express fusion constructs at endogenous levels, we expanded our toolbox to include a BirA* cassette which we knocked into the *Tet1* locus (Supplementary Figure S5G). In contrast to classical IP approaches, proximity-dependent protein labeling by the promiscuous biotin ligase, BirA* (BioID) (27), allows the characterization of the full microenvironment of a protein of interest independent of physical protein–protein interactions. This technique enabled us to pull down proteins within close proximity (~10 nm radius, (39)) of TET1 that were subsequently identified by LC-MS/MS (Figure 4C). We found nine proteins to be significantly enriched (40) upon addition of exogenous biotin to the culture medium of our *Tet1^{BirA*/BirA*}* mESC line, including SIN3A, a known interactor of TET1 (41) (Figure 4D and E). Interestingly, these proteins are associated with chromatin modification and organization (Figure 4F). This marks the first time that the BioID method has been used in mESCs and in a non-overexpression context with the BirA* ligase fused to the endogenous protein.

Using the MIN-tag strategy to study dynamic cellular processes

During early embryonic development, the epigenome undergoes massive rearrangements that are precisely regulated. Knockout of the major epigenetic factors is often embryonic lethal (38,42) and over-expression studies frequently fail to reflect the tight regulation of these proteins. Therefore, more flexible and delicate genetic manipulations

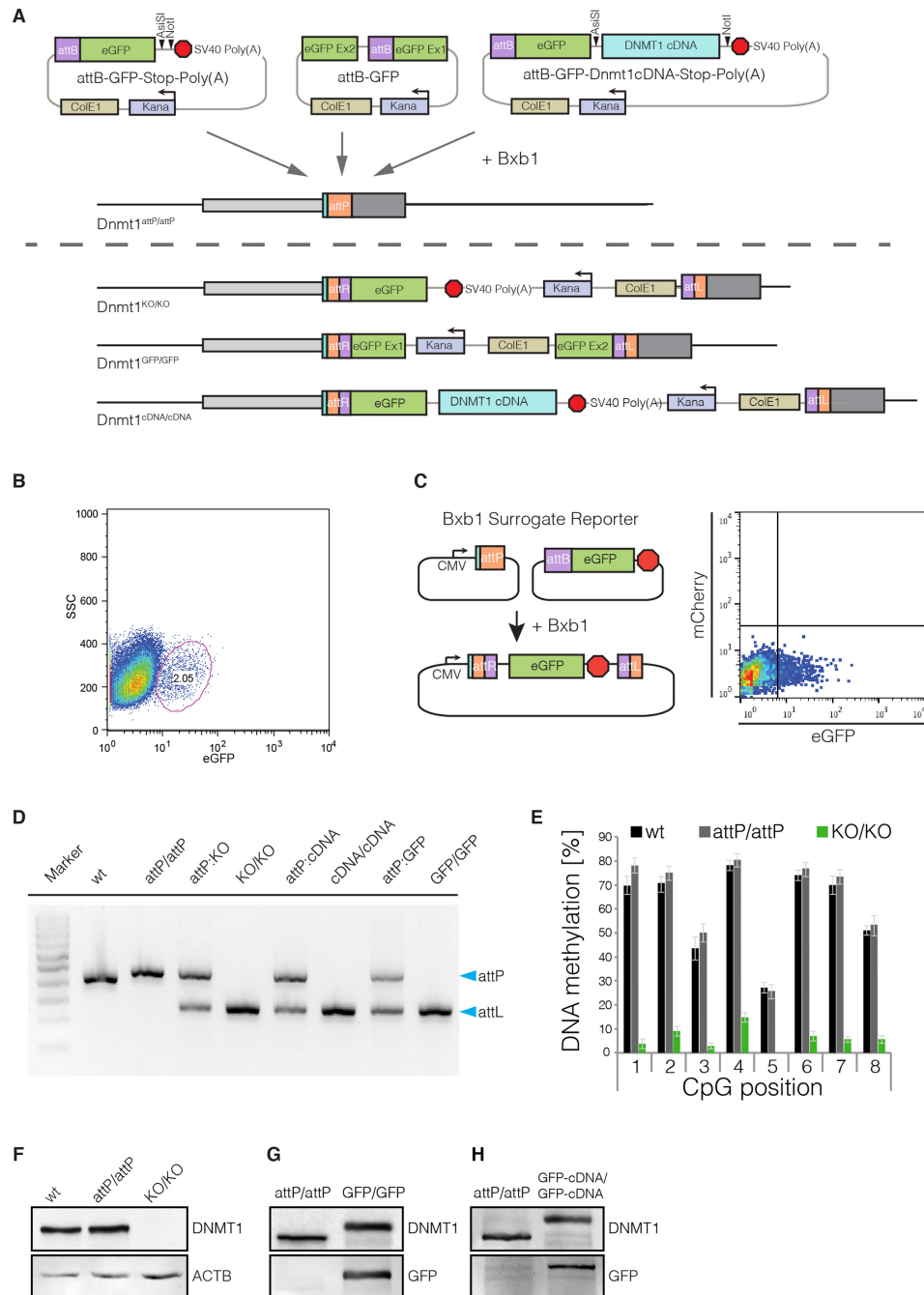


Figure 3. Bxb1-mediated insertion of functional cassettes into the *Dnmt1* locus. (A) Schematic outline of the strategy and vectors used to create knockout, GFP knockin and cDNA knockin functionalizations of the *Dnmt1*^{attP/attP} cell line. cDNAs can be cloned into the attB-GFP-Stop-Poly(A) vector using the 8-cutters AsiSI and NotI. (B) FACS plot depicting the gating and sorting of mESCs to enrich for cells positive for integration of the knockout cassette (2.05% of parent population) based on GFP expression. (C) The Bxb1 surrogate reporter consists of a constitutive CMV promoter followed by an attP site. If Bxb1 and attB donor plasmid containing GFP is present in the cell, recombination of the donor into the reporter leads to expression of GFP. The Bxb1 surrogate reporter can be used to enrich for successful recombination events by FACS. (D) Gel electrophoresis of the multiplex PCR for wt, *Dnmt1*^{attP/attP} (attP/attP), *Dnmt1*^{KO/KO} (KO/KO), *Dnmt1*^{cDNA/cDNA} (cDNA/cDNA) and *Dnmt1*^{GFP/GFP} (GFP/GFP) as well as 1:1 mixtures with *Dnmt1*^{attP/attP} genomic DNA, to control for amplification biases. Blue arrows indicate expected sizes of the non-recombined (attP) and recombined allele (attL). (E) DNA methylation levels at the major satellite repeats of *Dnmt1*^{KO/KO} cells compared to wt and *Dnmt1*^{attP/attP} cells. (F) Western blot analysis of DNMT1 expression levels in wt, *Dnmt1*^{attP/attP} and *Dnmt1*^{KO/KO} cells generated by Bxb1-mediated insertion of a knockout cassette. (G) Western blot analysis of DNMT1 and GFP expression in *Dnmt1*^{attP/attP} and homozygous GFP-knockin cells (*Dnmt1*^{GFP/GFP}) generated by Bxb1-mediated insertion of a knockout cassette. (H) Western blot analysis of DNMT1 and GFP expression in *Dnmt1*^{attP/attP} and *Dnmt1*^{cDNA/cDNA} cells expressing a GFP-Dnmt1 minigene from the endogenous promoter.

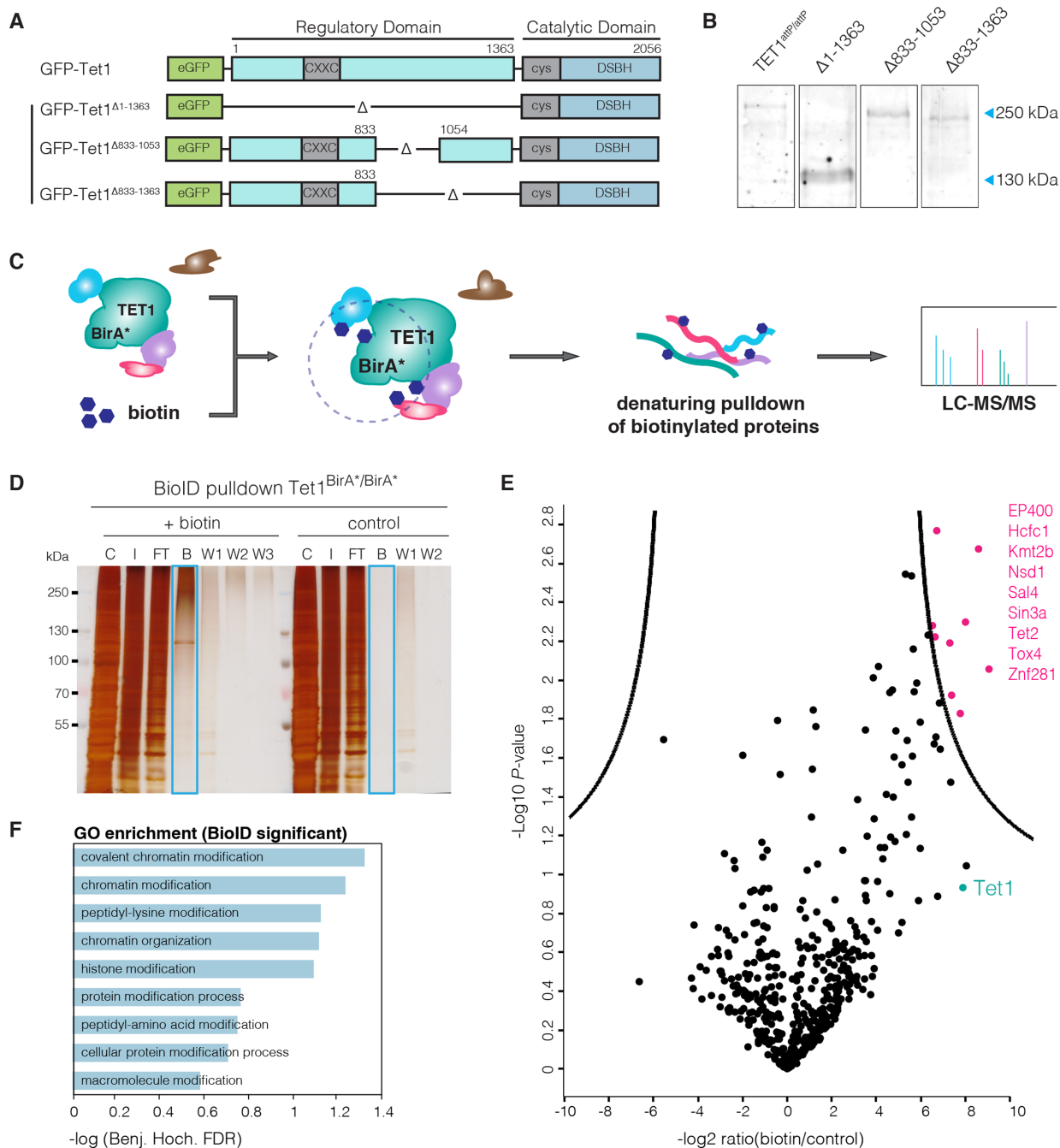


Figure 4. Study of TET1 regulation. (A) Schematic representation of the *Tet1* cDNA constructs used for Bxb1-mediated recombination into *Tet1^{attP/attP}* cells. (B) Western blot analysis of TET1 expression in *Tet1^{attP/attP}* cell line and its derivatives expressing GFP-TET1^{Δ1-1363} (Δ1-1363), GFP-TET1^{Δ833-1053} (Δ833-1053) and GFP-TET1^{Δ833-1363} (Δ833-1363). Note that fusion to GFP increases the MW of TET1 constructs by 29 kDa. (C) Schematic representation of the BioID approach as described by Roux *et al.* (27). (D) SDS-PAGE analysis of a BioID pulldown experiment using the Tet1^{BirA*/BirA*} cell line. Cells were cultured either without (control) or with 50 μM biotin (+biotin). C: Cytoplasm, I: Crude nuclei input, FT: Flowthrough, B: Bound, W1-W3: Wash. (E) Volcano plot of proteins identified in the streptavidin pulldown of the TET1-BioID experiment, quantified with the MaxQuant Label-Free-Quantification algorithm (32). The x-axis reflects the difference in protein abundance in the BioID pull-down compared to the negative control while the y-axis shows the logarithmized *P*-value of a student's *t*-test. Significantly enriched proteins are highlighted in pink (FDR = 0.01, S0 = 3, indicated by black line (40)). Experiments were performed in duplicates. (F) GO term enrichment of proteins identified as significant in BioID.

are needed to study the function of epigenetic factors *in vivo*. Here, we focus on the *de novo* DNA methyltransferase 3B (DNMT3B), one of the key factors during epiblast differentiation. While it has been shown that DNMT3B, in concert with DNMT3A and DNMT3L, is responsible for the global wave of *de novo* DNA methylation occurring during epiblast differentiation (42–44), little is known about its localization and protein kinetics during this developmental time period.

To address this question in a systematic fashion, we generated a homozygous GFP knockin cell line (*Dnmt3b^{GFP/GFP}*) from the *Dnmt3b^{attP/attP}* cell line by Bxb1-mediated recombination (Figure 5A and 6A). This allowed us to follow expression of DNMT3B under native regulatory conditions and to monitor its localization during the two-day transition from naive pluripotent ESCs to Epiblast-like cells (EpiLCs, (25)) using live cell imaging with high temporal resolution (1 image per hour).

At the naive pluripotent state, we observed very low expression levels of DNMT3B. Upon addition of differentiation medium, protein expression was strongly and uniformly upregulated reaching its maximum at 48–52 h (Figure 5B, Supplementary video 1). Overall, these findings were consistent with *Dnmt3b* mRNA levels in wt and *Dnmt3b^{attP/attP}* cells (Figure 5C). Interestingly, we observed a highly dynamic subnuclear distribution of DNMT3B during differentiation that can be classified into three patterns (Figure 5B). (i) In the first 14 h of differentiation, DNMT3B is expressed at low levels and no clear enrichment is visible. (ii) Between 14–40 h after initiation of differentiation, DNMT3B expression is upregulated and accumulates at constitutive heterochromatin of chromocenters (CCs). (iii) After 40 h of differentiation, DNMT3B is highly expressed and localization to CCs is diminished. The above-described patterns were not related to specific cell cycle stages, indicating a differentiation stage dependent localization of DNMT3B (Supplementary Figure S6A).

To investigate the specific chromatin distribution of DNMT3B during differentiation in more detail, we performed super-resolution 3D structured illumination microscopy (3D-SIM) with the anti-MIN antibody for protein visualization. DAPI and trimethylated lysine 4 of histone 3 (H3K4me3) were used as markers of heterochromatin and euchromatin (45), respectively. In agreement with the live cell imaging experiments, DNMT3B localizes at CCs, clusters of subcentromeric regions, at the 30 h time point and shows a broader distribution at 60 h after differentiation (Figure 5D). Interestingly, the higher resolution of 3D-SIM revealed an accumulation of the signal in facultative heterochromatin at perinuclear and perinucleolar regions at the 60 h time point (Figure 5D; right panel).

DNMT3B has been shown to be responsible for the methylation of major satellite DNA, a main constituent of CCs (42,46–47). As DNMT3B is enriched at CCs between 14–40 h of differentiation, we investigated whether DNMT3B is actively methylating these sequences during this period. Therefore, we performed fluorescence recovery after photobleaching (FRAP) of GFP-DNMT3B localized at CCs. Using our *Dnmt3b^{GFP/GFP}* cell line, we performed FRAP experiments at 35 h of differentiation. Using circular regions of interest (ROIs) that encompassed individual CCs, we monitored signal recovery for 10 min after

bleaching. We found that the signal exhibited a slow recovery rate ($t_{1/2} = 42$ s) and did not recover completely. As DNA methylation has been shown to have a slow turnover rate (48,49), this suggested the immobile fraction (~20%) of DNMT3B could be catalytically active at CCs (Figure 6B and D, Supplementary Table S4). To test this hypothesis, we performed FRAP experiments on cells treated with the DNA methyltransferase inhibitor 5-aza-2'-deoxycytidine (5-azadC), which irreversibly traps DNMTs at their site of action (50). We found that 5-azadC treated CCs exhibited a large immobile fraction (~80%) suggesting that DNMT3B is actively methylating CCs at this time point. However, we were surprised to find that ~20% of DNMT3B enzyme still remained mobile (Figure 6C). Considering the long 5-azadC treatment time of 12 h this suggested that a fraction of the enzyme never engaged in catalytic reactions. As our GFP cassette preserves endogenous splicing patterns, the GFP-DNMT3B fusions used in this study represent a mixture of different protein isoforms. This prompted us to investigate the contribution of *Dnmt3b* splicing isoforms to the observed FRAP kinetics.

For *Dnmt3b*, nine splicing isoforms, all originating from the same translational start site, have been described (51). Besides the catalytically active isoform DNMT3B1, DNMT3B6 has been shown to be highly expressed in ESCs. This isoform is produced by alternative splicing, skipping exons 23 and 24, resulting in a protein that lacks several highly conserved motifs within the catalytic domain and has therefore been suggested to be inactive (52).

To dissect the contributions of DNMT3B1 and DNMT3B6 to the observed FRAP kinetics of *Dnmt3b^{GFP/GFP}* cells, we generated a cell line expressing fluorescent fusions of each isoform. For this, we produced cDNA knockin constructs in which DNMT3B1 was fused to a red fluorescent protein mCherry (mCh) and DNMT3B6 was fused to GFP. To facilitate the generation of knockin cell lines expressing each isoform from one allele we equipped the *Dnmt3b1* and *Dnmt3b6* constructs with a Neomycin and Puromycin resistance cassette, respectively. We successfully established a cell line that simultaneously expressed mCh-DNMT3B1 and GFP-DNMT3B6, both under the control of the endogenous *Dnmt3b* promoter (Figure 6A, Supplementary Figure S6B), allowing us to directly compare the FRAP kinetics of DNMT3B1 and DNMT3B6 within the same cell. In the absence of 5-azadC, GFP-DNMT3B6 exhibited a fast ($t_{1/2} = 5$ s) and complete recovery while mCh-DNMT3B1 recovered slower ($t_{1/2} = 95$ s) (Figure 6B, Supplementary Table S4).

Intriguingly, FRAP kinetics of DNMT3B6 were not influenced by the presence of 5-azadC, supporting that it is catalytically inactive. In contrast, DNMT3B1 was completely immobilized by addition of 5-azadC exhibiting virtually no recovery after photobleaching (Figure 6C and E).

Taken together, our MIN-tag strategy enabled us to show that DNMT3B exhibits a dynamic localization to distinct chromatin regions during epiblast differentiation. Super-resolution micrographs of cells stained with anti-MIN antibodies at different time points of epiblast differentiation hint towards progression of *de novo* DNA methylation in a hierarchical fashion starting at constitutive (CCs) and progressing towards facultative (perinuclear/perinucleolar)

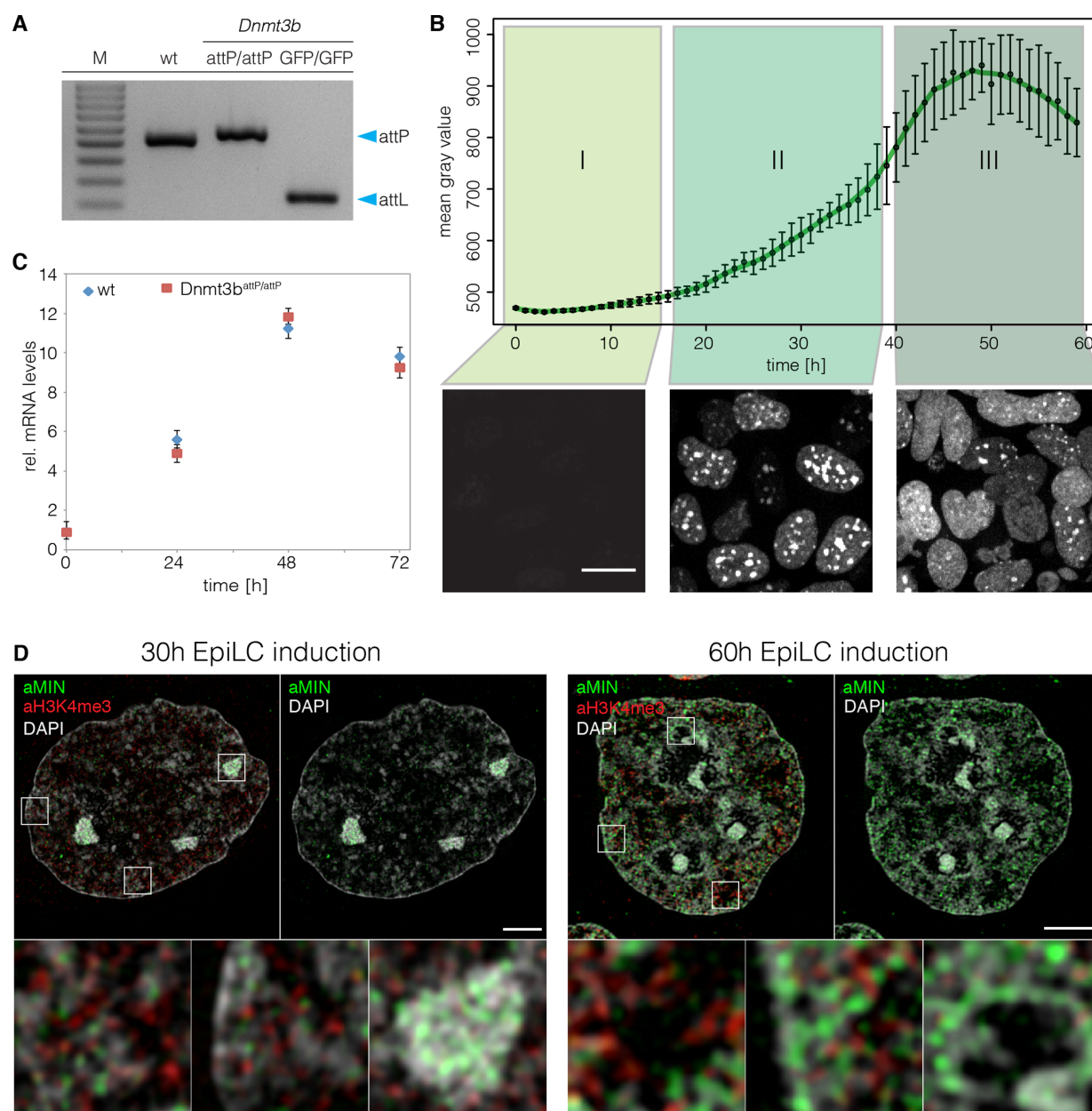


Figure 5. Spatio-temporal dynamics of DNMT3B during epiblast differentiation. (A) Gel electrophoresis of the multiplex screening PCR for wt, *Dnmt3b*^{attP/attP} and *Dnmt3b*^{GFP/GFP}. Blue arrows indicate expected sizes of the non-recombined (attP) and recombined allele (attL). (B) Evaluation of GFP signals during live cell imaging of *Dnmt3b*^{GFP/GFP} cells. The graph depicts mean gray values of nuclear GFP signals. Error bars represent standard deviations ($n > 81$). Lower panels show Z-projections of *Dnmt3b*^{GFP/GFP} cells representative of the indicated time frame. Scale bar represents 10 μ m. (C) Quantitative real-time PCR of *Dnmt3b* mRNA levels in wt and *Dnmt3b*^{attP/attP} cells during epiblast differentiation. (D) 3D-SIM nuclear mid-sections of anti-MIN (green) and anti-H3K4me3 (red) antibody distributions 30 and 60 h after induction of EpiLC differentiation combined with DAPI counterstaining (gray) in *Dnmt3b*^{attP/attP} cells. Lower panels represent 7 \times magnifications of selected boxed regions. Scale bars represent 3 μ m and 500 nm in insets.

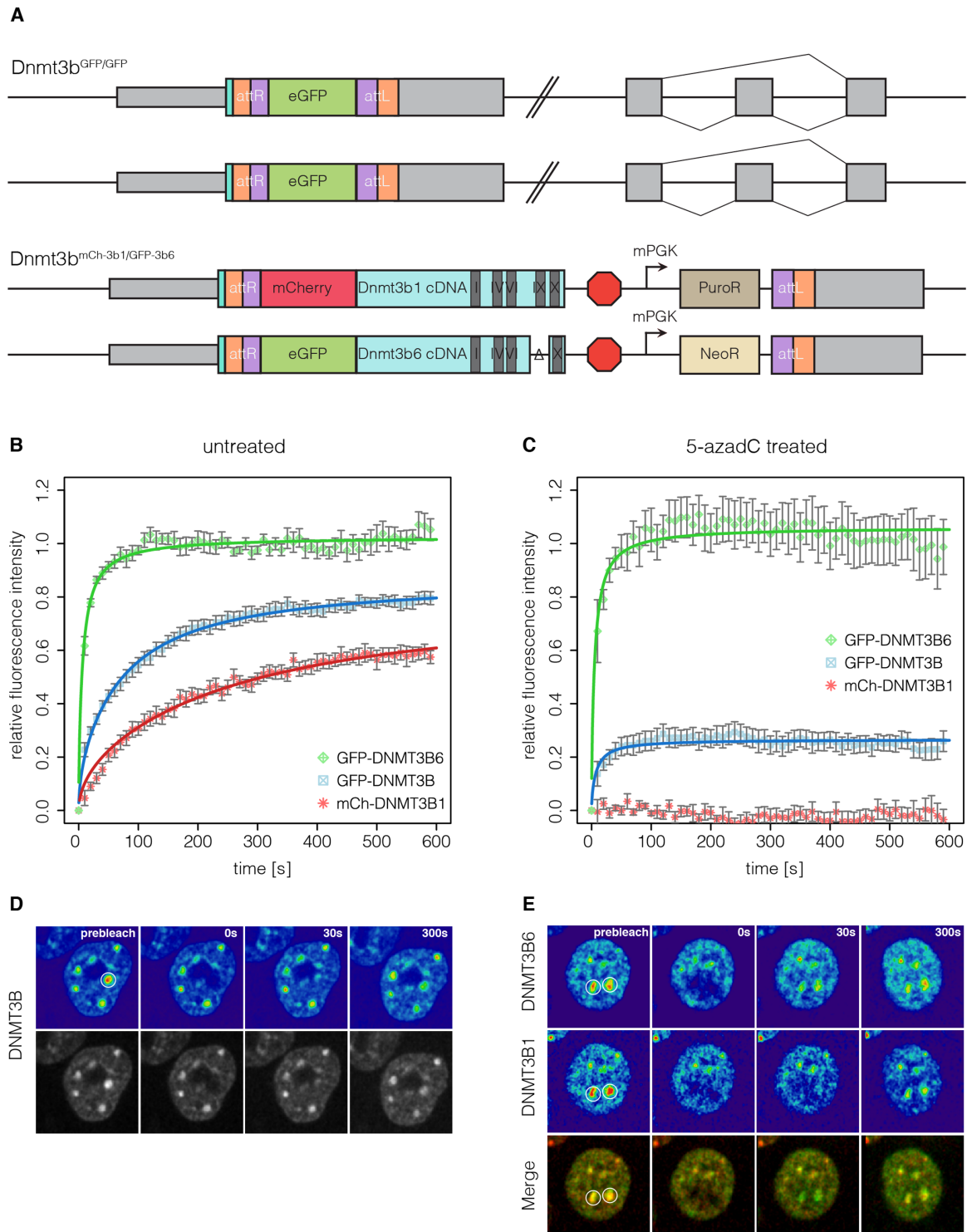


Figure 6. Protein dynamics of DNMT3B and its isoforms during epiblast differentiation. (A) Schematic representation of the *Dnmt3b* genomic loci in the *Dnmt3b*^{GFP/GFP} and the *Dnmt3b*^{mCh-3b1/GFP-3b6} cell lines. (B) Quantitative evaluation of FRAP experiments (average of 11–14 cells) comparing GFP-DNMT3B with GFP-DNMT3B6 and mCh-DNMT3B1 in *Dnmt3b*^{GFP/GFP} and the *Dnmt3b*^{mCh-3b1/GFP-3b6} cell lines differentiated for 35 h. Error bars represent standard error of the mean. (C) Quantitative evaluation of FRAP experiments (average of 10–12 cells) as in (B) with cells treated with 5-azadC 12 h before imaging. (D and E) Representative images of FRAP experiments performed in (B) and (C), respectively. White circles indicate the bleach ROI with a diameter of 2 μ m.

heterochromatin. Finally, FRAP experiments revealed that the two isoforms DNMT3B1 and DNMT3B6 exhibit dramatically different DNA binding kinetics.

DISCUSSION

Recent advances in genome engineering technology, based on TALEN and CRISPR/Cas systems, have greatly facilitated the process of manipulating genetic information. Platforms have been established that allow genome-wide gene disruption screenings for factors involved in any biological process (20,53–54). While these methods provide valuable information about the genes and pathways involved, in-depth analysis of target genes is needed to understand their function. This, in turn, requires the implementation of various genetic, cell biological and biochemical techniques. To gain meaningful insights into gene function, these techniques have to be applied under physiological conditions requiring extensive and complex genetic manipulations. Although modern genome engineering tools have made such manipulations possible, a more efficient and universal approach would be highly desirable to implement the above-mentioned techniques in a systematic manner.

The MIN-tag strategy offers a new means of rapid, efficient, yet flexible genetic manipulation of target loci. We show that CRISPR/Cas assisted insertion of the MIN-tag can be performed efficiently with short homology donors. Several studies have shown that CRISPR/Cas mediated gene targeting is associated with a significant risk of off-target cleavage, which can result in indel (insertions or deletions) formation due to NHEJ (5–7,55–56). The MIN-tag strategy requires a single nuclease assisted gene editing event, thereby keeping the likelihood of off-target effects at a minimum. Further modifications are then performed using Bxb1-mediated recombination. In contrast to the phiC31 integrase, Bxb1 has been shown to be highly specific with virtually no unwanted genomic insertions at pseudo *attP* sites (8–9,57–58). Once a MIN-tagged cell line is established, in-frame fusion of the MIN-tag to the target gene also results in the expression of a novel epitope tag. We show that this epitope tag can be detected by a highly specific antibody, which can be used to screen for positive clones, perform co-immunoprecipitation (co-IP) experiments, as well as conventional and super resolution microscopy.

Using Bxb1 and the MIN-tag toolbox, a MIN-tagged entry cell line can be used to generate multiple isogenic derivatives within 2–3 weeks (Figure 7), without the risk of introducing off-target effects. Our collection of vectors for Bxb1 mediated recombination currently contains over 80 different plasmids (Supplementary Table S5). These prefabricated functional cassettes constitute an expandable toolbox for the simple and flexible genetic alteration of any tagged loci, without the need of locus-specific homology.

Using our stop cassette, we show that the MIN-tag strategy can be used to reliably achieve genetically defined gene disruption of MIN-tagged genes. Harboring a Poly(A) signal, insertion of this cassette efficiently eliminates target gene expression with the added advantage of precluding unwanted downstream initiation. As fluorescent protein reporters are commonly used to study spatio-temporal dy-

namics and protein kinetics in living cells, we generated a GFP knockin construct (*attB*-GFP) for Bxb1-mediated integration. GFP knockin cell lines made with this construct retain not only their endogenous expression levels but also their endogenous splicing pattern. Similarly, a BirA* cassette can be introduced at any MIN-tagged locus to allow for proximity-dependent labeling of the microenvironment of a given protein.

Understanding protein function often necessitates the systematic alteration of individual domains through mutations as well as deletions. Equipped with a fluorescent protein and strategic cloning sites, our cDNA knockin cassette is especially tailored for simple and expedient insertion of user-defined cDNAs. PCR-based approaches can be used to easily alter the coding sequence and quickly produce a library of gene specific cDNA mutants. These can then be inserted into target loci by Bxb1-mediated recombination, completely replacing expression of the wt gene while retaining endogenous control. While this strategy does not directly introduce the mutations into the gene locus, it offers a means of inserting and testing multiple mutant constructs in a short time frame without the need to design and perform additional nuclease-assisted targetings. This feature can be used to gain insights into the functional implications of the rapidly growing number of mutations found in cancer and disease. Likewise, the generation of large deletion mutants is easily accomplished facilitating the investigation of protein domain function and interaction mapping. This eliminates the need for excising large genomic regions or cloning long site-specific homology donors.

Obviously, the above mentioned plasmids by no means represent the extent of all possible functional cassettes. For example, MIN-tag toolbox modules allowing inducible protein stabilization or localization (59,60) as well as enzymatic labeling of DNA binding sites (DamID (61)) would greatly assist the elucidation of protein function and protein-chromatin interactions, respectively.

Employing our strategy in mESCs, we inserted the MIN-tag into the genes coding for all mammalian DNA modifying enzymes and a cofactor (*Dnmt1*, *Dnmt3a*, *Dnmt3b*, *Tet1*, *Tet2*, *Tet3* and *Uhrf1*). These MIN-tagged cell lines as well as their functional derivatives (Supplementary Table S3) constitute a valuable resource to investigate the role of these proteins during fundamental processes such as pluripotency, cellular reprogramming, embryonic development and disease.

One gold standard method to study protein–protein interactions is co-IP. However, chromatin- or membrane-bound proteins are often barely soluble and consequently difficult to investigate by this approach. Making use of our BirA* cassette, we investigated factors in the microenvironment of TET1, a dioxygenase that oxidizes DNA at methylated cytosines (62). Besides the known interactor SIN3A, we identify eight other proteins in proximity to TET1 that are involved in chromatin modification and organization, including the closely related TET2. This is in accordance with the findings by Costa *et al.* (63) that TET1 and TET2 have partially overlapping target sites. In conclusion, integration of the BirA* cassette into the endogenous locus is a perfectly suited method to study dynamic protein–protein interactions.

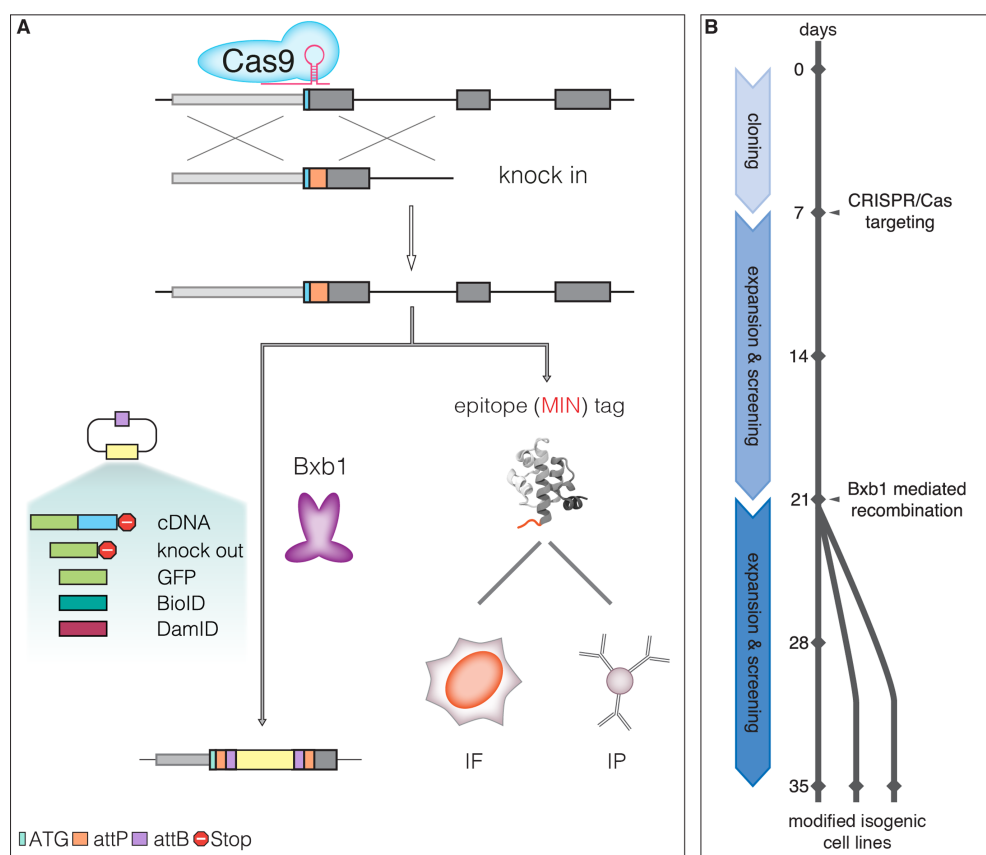


Figure 7. The MIN-tag strategy. **(A)** Schematic outline of the genome engineering strategy. Small homology donors are used to insert serine integrase (*attP*) sites in-frame after the ATG codon of target genes via CRISPR/Cas assisted HR. The *attP* site is translated as a novel epitope tag suitable for IF and IP with the specific monoclonal antibody. The *attP* site is also recognized by the serine integrase Bxb1 and used for specific and directional integration of *attB*-carrying functional cassettes into the tagged gene locus. All derivatives are subjected to their endogenous gene regulation ensuring that subsequent studies are performed at physiological expression levels. **(B)** Timeline for generation of MIN-tagged genes and subsequent modification by Bxb1-mediated recombination. MIN-tagged cell lines can be generated within 2–3 weeks. These cell lines can then be modified within another 2–3 weeks to generate multiple isogenic cell lines with different functional modifications.

We also applied the MIN-tag strategy to study the *de novo* DNA methyltransferase DNMT3B during the transition from naive pluripotent ESCs to primed EpiLCs, a period of dramatic epigenetic change. While distinct patterns have been described for ESCs and somatic cells (46,64), the subnuclear distribution of DNMT3B during differentiation remains largely unknown. We discovered that DNMT3B exhibits a highly dynamic subnuclear distribution during epiblast differentiation. Our observations suggest that the global wave of *de novo* DNA methylation during epiblast differentiation follows a distinct spatio-temporal order, initiating at constitutive pericentromeric heterochromatin followed by transition to facultative heterochromatin.

Exploiting the unique possibilities of our MIN-tag strategy, we furthermore generated a cell line simultaneously expressing differentially tagged splicing isoforms of DNMT3B from different alleles. This approach revealed that the major catalytically active isoform DNMT3B1 was completely immobilized at chromocenters after 5-azadC treatment, while the FRAP kinetics of DNMT3B6 were not affected. This, to our knowledge, is the first time that FRAP

has been performed on different isoforms of a protein at endogenous expression levels in the same cell.

While this study was performed using mouse ESCs, our strategy can be applied to any cell type as long as no Bxb1 *attP* site is present in the respective genomes. The human genome is free of this entry site and introduction of the MIN-tag into cell lines such as human induced pluripotent stem cells should greatly facilitate the generation of clinically relevant disease models. Moreover, MIN-tagged mESCs could be used in blastocyst injections to generate MIN-tagged mice. Different tissues and cells could not only be used for Bxb1-mediated genetic manipulation *in vitro*, free of the limitation posed by inefficient endogenous homologous recombination, but also to study tissue specific protein regulation with the MIN-tag antibody. Furthermore, widely used cell biological model systems such as HeLa and U2OS cells as well as model organisms such as *Caenorhabditis elegans* or *Drosophila* could benefit from the versatility and efficiency of our approach.

In summary, with our combined genome engineering approach, a plethora of functional derivatives can be gener-

ated from one entry line with high efficiency. To simplify the distribution of MIN-tagged cell lines and the MIN-tag toolbox as well as to assist with the design of targeting strategies, we have developed a web-tool that is accessible at http://human.bio.lmu.de/_webtools/MINtool/. As entry lines can be shared and the genetic toolbox easily expanded with new functional modules, the MIN-tag strategy represents a dynamic flexible open platform and facilitates systematic functional studies with direct biological comparability.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Ina Poser and Tony Hyman (Max Planck Institute of Molecular Biology & Genetics, Dresden) for providing R6K-NFLAP construct, George Church (Harvard Medical School, Boston) for providing the Cas9 expression construct, Kerry Tucker (Ruprecht-Karls-University, Heidelberg) for providing wt ESCs and Pawel Pelczar (University of Zürich) for providing pCAG-NLS-HA-Bxb1. Furthermore, we thank Robert Engelmeier and Michael Soutschek for help with cell line generation, IF and antibody characterization. We also thank Andy Spiegl, Gregor Jessberger and Jan Langkabel for help with website development. CBM gratefully acknowledges the Life Science Munich Graduate School. ES, CB and KT gratefully acknowledge the International Max Planck Research School for Molecular and Cellular Life Sciences.

FUNDING

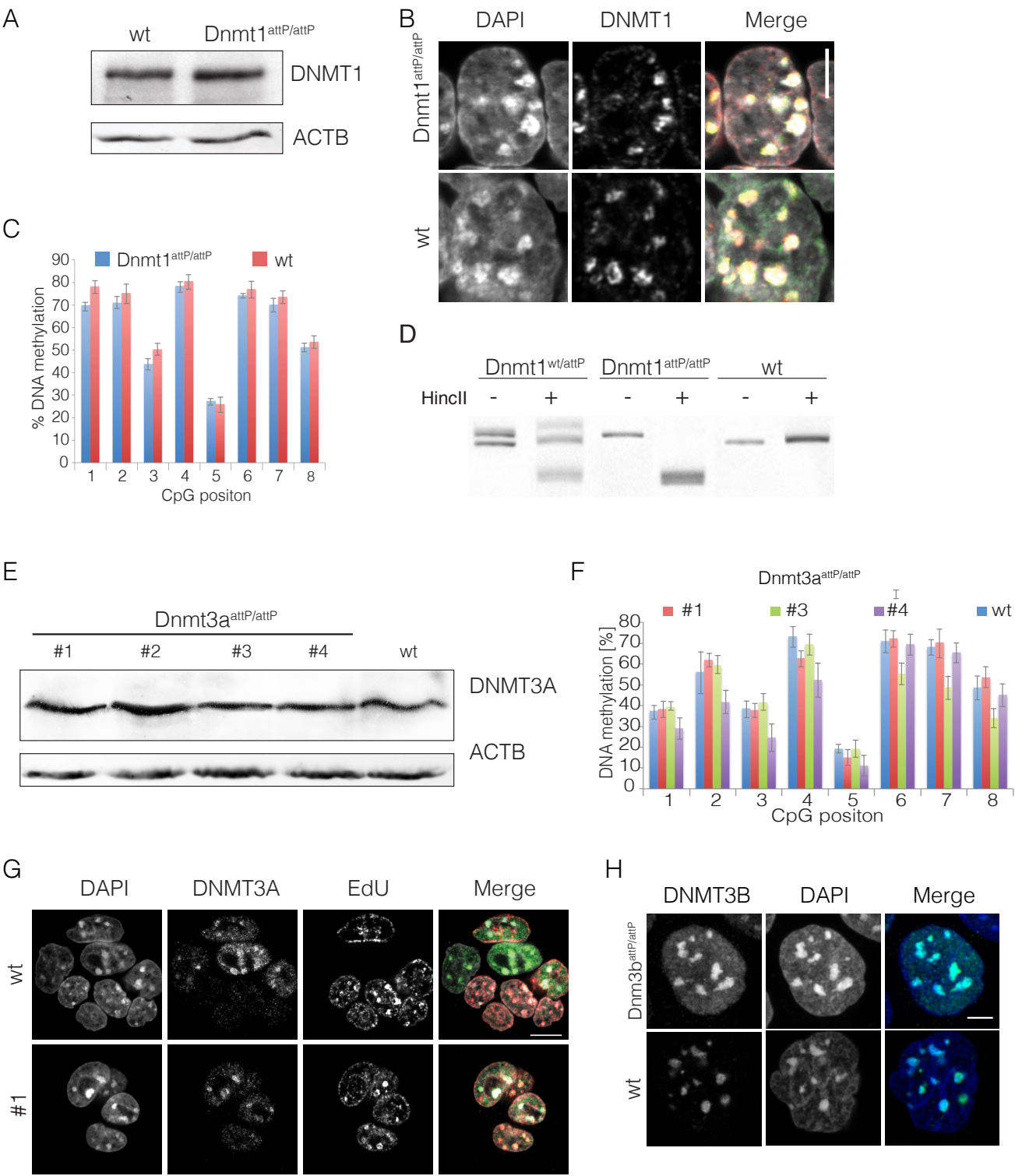
Deutsche Forschungsgemeinschaft [SFB 1064 to H.L., E.K. and GRK 1721 to H.L., F.H.]; Funding for open access charge: Deutsche Forschungsgemeinschaft [Collaborative Research Center SFB 1064/A17; GRK 1721].
Conflict of interest statement. None declared.

REFERENCES

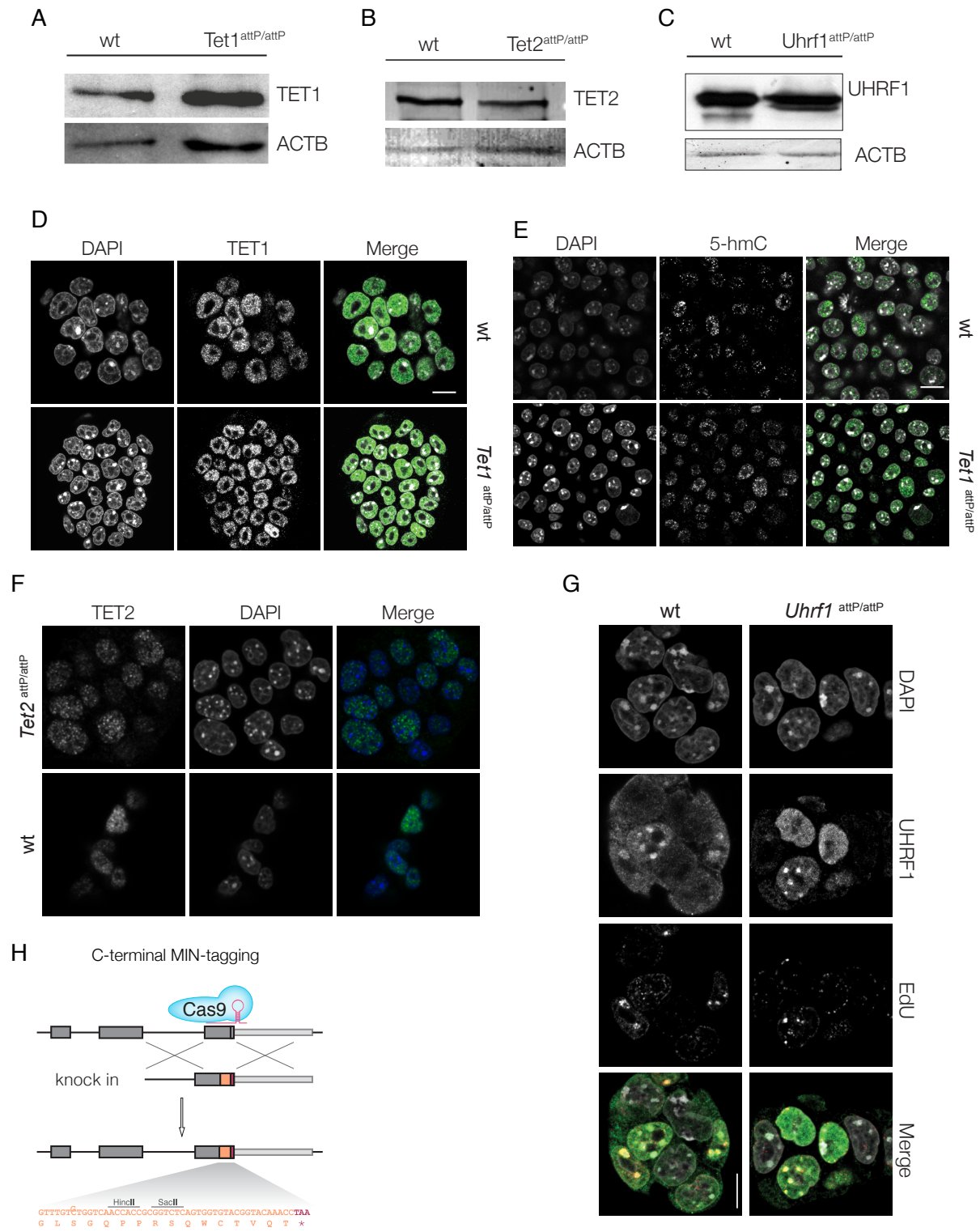
- Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A. *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science*, **339**, 819–823.
- Haurwitz, R.E., Jinek, M., Wiedenheft, B., Zhou, K. and Doudna, J.A. (2010) Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science*, **329**, 1355–1358.
- Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E. and Church, G.M. (2013) RNA-guided human genome engineering via Cas9. *Science*, **339**, 823–826.
- Sampson, T.R., Saroj, S.D., Llewellyn, A.C., Tzeng, Y.-L. and Weiss, D.S. (2013) A CRISPR/Cas system mediates bacterial innate immune evasion and virulence. *Nature*, **497**, 254–257.
- Kuscu, C., Arslan, S., Singh, R., Thorpe, J. and Adli, M. (2014) Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nat. Biotechnol.*, **32**, 677–683.
- Wang, X., Wang, Y., Wu, X., Wang, J., Wang, Y., Qiu, Z., Chang, T., Huang, H., Lin, R.-J. and Yee, J.-K. (2015) Unbiased detection of off-target cleavage by CRISPR-Cas9 and TALENs using integrase-defective lentiviral vectors. *Nat. Biotechnol.*, **33**, 175–178.
- Wu, X., Scott, D.A., Kriz, A.J., Chiu, A.C., Hsu, P.D., Dadon, D.B., Cheng, A.W., Trevino, A.E., Konermann, S., Chen, S. *et al.* (2014) Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat. Biotechnol.*, **32**, 670–676.
- Xu, Z., Thomas, L., Davies, B., Chalmers, R., Smith, M. and Brown, W. (2013) Accuracy and efficiency define Bxb1 integrase as the best of fifteen candidate serine recombinases for the integration of DNA into the human genome. *BMC Biotechnol.*, **13**, 87–87.
- Brown, W.R.A., Lee, N.C.O., Xu, Z. and Smith, M.C.M. (2011) Serine recombinases as tools for genome engineering. *Methods*, **53**, 372–379.
- Bonnet, J., Subsoontorn, P. and Endy, D. (2012) Rewritable digital data storage in live cells via engineered control of recombination directionality. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 8884–8889.
- Huang, J., Ghosh, P., Hatfull, G.F. and Hong, Y. (2011) Successive and targeted DNA integrations in the *Drosophila* genome by Bxb1 and phiC31 integrases. *Genetics*, **189**, 391–395.
- Zhu, F., Gamboa, M., Farruggio, A.P., Hippenmeyer, S., Tasic, B., Schule, B., Chen-Tsai, Y. and Calos, M.P. (2014) DICE, an efficient system for iterative genomic editing in human pluripotent stem cells. *Nucleic Acids Res.*, **42**, e34.
- Citterio, E., Papait, R., Nicassio, F., Vecchi, M., Gomiero, P., Mantovani, R., Di Fiore, P.P. and Bonapace, I.M. (2004) Np95 is a histone-binding protein endowed with ubiquitin ligase activity. *Mol. Cell. Biol.*, **24**, 2526–2535.
- Bauer, C., Gobel, K., Nagaraj, N., Colantuoni, C., Wang, M., Muller, U., Kremmer, E., Rottach, A. and Leonhardt, H. (2015) Phosphorylation of TET proteins is regulated via O-GlcNAcylation by the O-linked N-acetylglucosamine transferase (OGT). *J. Biol. Chem.*, **290**, 4801–4812.
- Schneider, K., Fuchs, C., Dobay, A., Rottach, A., Qin, W., Wolf, P., Álvarez-Castro, J.M., Nalaskowski, M.M., Kremmer, E., Schmid, V. *et al.* (2013) Dissection of cell cycle-dependent dynamics of Dnmt1 by FRAP and diffusion-coupled modeling. *Nucleic Acids Res.*, **41**, 4860–4876.
- Solovei, I. and Cremer, M. (2010) 3D-FISH on cultured cells combined with immunostaining. *Methods Mol. Biol.*, **659**, 117–126.
- Meilinger, D., Fellinger, K., Bultmann, S., Rothbauer, U., Bonapace, I.M., Klinkert, W.E.F., Spada, F. and Leonhardt, H. (2009) Np95 interacts with de novo DNA methyltransferases, Dnmt3a and Dnmt3b, and mediates epigenetic silencing of the viral CMV promoter in embryonic stem cells. *EMBO Rep.*, **10**, 1259–1264.
- Niwa, H., Yamamura, K. and Miyazaki, J. (1991) Efficient selection for high-expression transfectants with a novel eukaryotic vector. *Gene*, **108**, 193–199.
- Poser, I., Sarov, M., Hutchins, J.R.A., Hériché, J.-K., Toyoda, Y., Pozniakovskiy, A., Weigl, D., Nitzsche, A., Hegemann, B., Bird, A.W. *et al.* (2008) BAC TransgeneOmics: a high-throughput method for exploration of protein function in mammals. *Nat. Methods*, **5**, 409–415.
- Schmid-Burgk, J.L., Schmidt, T., Kaiser, V., Honing, K. and Hornung, V. (2013) A ligation-independent cloning technique for high-throughput assembly of transcription activator-like effector genes. *Nat. Biotechnol.*, **31**, 76–81.
- Masui, S., Shimosato, D., Toyooka, Y., Yagi, R., Takahashi, K. and Niwa, H. (2005) An efficient system to establish multiple embryonic stem cell lines carrying an inducible expression unit. *Nucleic Acids Res.*, **33**, e43.
- Easwaran, H.P., Schermelleh, L., Leonhardt, H. and Cardoso, M.C. (2004) Replication-independent chromatin loading of Dnmt1 during G2 and M phases. *EMBO Rep.*, **5**, 1181–1186.
- Frauer, C., Rottach, A., Meilinger, D., Bultmann, S., Fellinger, K., Hasenoder, S., Wang, M., Qin, W., Soding, J., Spada, F. *et al.* (2011) Different binding properties and function of CXXC zinc finger domains in Dnmt1 and Tet1. *PLoS One*, **6**, e16627.
- Schermelleh, L., Haemmer, A., Spada, F., Rosing, N., Meilinger, D., Rothbauer, U., Cardoso, M.C. and Leonhardt, H. (2007) Dynamics of Dnmt1 interaction with the replication machinery and its role in postreplicative maintenance of DNA methylation. *Nucleic Acids Res.*, **35**, 4301–4312.
- Hayashi, K. and Saitou, M. (2013) Generation of eggs from mouse embryonic stem cells and induced pluripotent stem cells. *Nat. Protoc.*, **8**, 1513–1524.
- Hermann, M., Stillhard, P., Wildner, H., Seruggia, D., Kapp, V., Sanchez-Iranzo, H., Mercader, N., Montoliu, L., Zeilhofer, H.U. and Pelczar, P. (2014) Binary recombinase systems for high-resolution conditional mutagenesis. *Nucleic Acids Res.*, **42**, 3894–3907.

27. Roux,K.J., Kim,D.I., Raida,M. and Burke,B. (2012) A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J. Cell Biol.*, **196**, 801–810.
28. Baymaz,H.I., Spruijt,C.G. and Vermeulen,M. (2014) Identifying nuclear protein-protein interactions using GFP affinity purification and SILAC-based quantitative mass spectrometry. *Methods Mol. Biol.*, **1188**, 207–226.
29. Blum,H., Beier,H. and Gross,H.J. (1987) Improved silver staining of plant-proteins, RNA and DNA in polyacrylamide gels. *Electrophoresis*, **8**, 93–99.
30. Rappsilber,J., Mann,M. and Ishihama,Y. (2007) Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.*, **2**, 1896–1906.
31. Cox,J. and Mann,M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.*, **26**, 1367–1372.
32. Cox,J., Hein,M.Y., Luber,C.A., Paron,I., Nagaraj,N. and Mann,M. (2014) Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteomics*, **13**, 2513–2526.
33. UniProt Consortium. (2015) UniProt: a hub for protein information. *Nucleic Acids Res.*, **43**, D204–D212.
34. Eden,E., Navon,R., Steinfeld,I., Lipson,D. and Yakhini,Z. (2009) GOrrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics*, **10**, 48–48.
35. Kim,H., Um,E., Cho,S.-R., Jung,C., Kim,H. and Kim,J.-S. (2011) Surrogate reporters for enrichment of cells with nuclease-induced mutations. *Nat. Methods*, **8**, 941–943.
36. Geiman,T.M., Sankpal,U.T., Robertson,A.K., Chen,Y., Mazumdar,M., Heale,J.T., Schmiesing,J.A., Kim,W., Yokomori,K., Zhao,Y. *et al.* (2004) Isolation and characterization of a novel DNA methyltransferase complex linking DNMT3B with components of the mitotic chromosome condensation machinery. *Nucleic Acids Res.*, **32**, 2716–2729.
37. Lehnertz,B., Ueda,Y., Derijck,A.A.H.A., Braunschweig,U., Perez-Burgos,L., Kubicek,S., Chen,T., Li,E., Jenuwein,T. and Peters,A.H.F.M. (2003) Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. *Curr. Biol.*, **13**, 1192–1200.
38. Li,E., Bestor,T.H. and Jaenisch,R. (1992) Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell*, **69**, 915–926.
39. Kim,D.I., Birendra,K.C., Zhu,W., Motamedchaboki,K., Doye,V. and Roux,K.J. (2014) Probing nuclear pore complex architecture with proximity-dependent biotinylation. *Proc. Natl. Acad. Sci. U.S.A.*, **111**, 2453–2461.
40. Tusher,V.G., Tibshirani,R. and Chu,G. (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 5116–5121.
41. Williams,K., Christensen,J., Pedersen,M.T., Johansen,J.V., Cloos,P.A.C., Rappsilber,J. and Helin,K. (2011) TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature*, **473**, 343–348.
42. Okano,M., Bell,D.W., Haber,D.A. and Li,E. (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*, **99**, 247–257.
43. Hata,K., Okano,M., Lei,H. and Li,E. (2002) Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. *Development*, **129**, 1983–1993.
44. Okano,M., Xie,S. and Li,E. (1998) Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nat. Genet.*, **19**, 219–220.
45. Bernstein,B.E., Kamal,M., Lindblad-Toh,K., Bekiranov,S., Bailey,D.K., Huebert,D.J., McMahon,S., Karlsson,E.K., Kulbokas,E.J., Gingeras,T.R. *et al.* (2005) Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell*, **120**, 169–181.
46. Bachman,K.E., Rountree,M.R. and Baylin,S.B. (2001) Dnmt3a and Dnmt3b are transcriptional repressors that exhibit unique localization properties to heterochromatin. *J. Biol. Chem.*, **276**, 32282–32287.
47. Xu,G.L., Bestor,T.H., Bourc'his,D., Hsieh,C.L., Tommerup,N., Bugge,M., Hulten,M., Qu,X., Russo,J.J. and Viegas-Pequignot,E. (1999) Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature*, **402**, 187–191.
48. Emperle,M., Rajavelu,A., Reinhardt,R., Jurkowska,R.Z. and Jeltsch,A. (2014) Cooperative DNA binding and protein/DNA fiber formation increases the activity of the Dnmt3a DNA methyltransferase. *J. Biol. Chem.*, **289**, 29602–29613.
49. Gowher,H. and Jeltsch,A. (2001) Enzymatic properties of recombinant Dnmt3a DNA methyltransferase from mouse: the enzyme modifies DNA in a non-processive manner and also methylates non-CpG [correction of non-CpA] sites. *J. Mol. Biol.*, **309**, 1201–1208.
50. Schermelleh,L., Spada,F., Easwaran,H.P., Zolghadr,K., Margot,J.B., Cardoso,M.C. and Leonhardt,H. (2005) Trapped in action: direct visualization of DNA methyltransferase activity in living cells. *Nat. Methods*, **2**, 751–756.
51. Pruitt,K.D., Brown,G.R., Hiatt,S.M., Thibaud-Nissen,F., Astashyn,A., Ermolaeva,O., Farrell,C.M., Hart,J., Landrum,M.J., McGarvey,K.M. *et al.* (2014) RefSeq: an update on mammalian reference sequences. *Nucleic Acids Res.*, **42**, 756–763.
52. Weisenberger,D.J., Velicescu,M., Cheng,J.C., Gonzales,F.A., Liang,G. and Jones,P.A. (2004) Role of the DNA methyltransferase variant DNMT3b3 in DNA methylation. *Mol. Cancer Res.*, **2**, 62–72.
53. Shalem,O., Sanjana,N.E., Hartenian,E., Shi,X., Scott,D.A., Mikkelsen,T.S., Heckl,D., Ebert,B.L., Root,D.E., Doench,J.G. *et al.* (2014) Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science*, **343**, 84–87.
54. Wang,T., Wei,J.J., Sabatini,D.M. and Lander,E.S. (2014) Genetic screens in human cells using the CRISPR-Cas9 system. *Science*, **343**, 80–84.
55. Lin,Y., Cradick,T.J., Brown,M.T., Deshmukh,H., Ranjan,P., Sarode,N., Wile,B.M., Vertino,P.M., Stewart,F.J. and Bao,G. (2014) CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences. *Nucleic Acids Res.*, **42**, 7473–7485.
56. Tan,E.P., Li,Y., Velasco-Herrera Mdel,C., Yusa,K. and Bradley,A. (2015) Off-target assessment of CRISPR-Cas9 guiding RNAs in human iPS and mouse ES cells. *Genesis*, **53**, 225–236.
57. Russell,J.P., Chang,D.W., Tretiakova,A. and Padidam,M. (2006) Phage Bxb1 integrase mediates highly efficient site-specific recombination in mammalian cells. *Biotechniques*, **40**, 460–460.
58. Keravala,A., Groth,A.C., Jarrahan,S., Thyagarajan,B., Hoyt,J.J., Kirby,P.J. and Calos,M.P. (2006) A diversity of serine phage integrases mediate site-specific recombination in mammalian cells. *Mol. Genet. Genomics*, **276**, 135–146.
59. Banaszynski,L.A., Chen,L.-C., Maynard-Smith,L.A., Ooi,A.G.L. and Wandless,T.J. (2006) A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules. *Cell*, **126**, 995–1004.
60. Kennedy,M.J., Hughes,R.M., Peteya,L.A., Schwartz,J.W., Ehlers,M.D. and Tucker,C.L. (2010) Rapid blue-light-mediated induction of protein interactions in living cells. *Nat. Methods*, **7**, 973–975.
61. van Steensel,B., Delrow,J. and Henikoff,S. (2001) Chromatin profiling using targeted DNA adenine methyltransferase. *Nat. Genet.*, **27**, 304–308.
62. Tahiliani,M., Koh,K.P., Shen,Y., Pastor,W.A., Bandukwala,H., Brudno,Y., Agarwal,S., Iyer,L.M., Liu,D.R., Aravind,L. *et al.* (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*, **324**, 930–935.
63. Costa,Y., Ding,J., Theunissen,T.W., Faiola,F., Hore,T.A., Shliha,P.V., Fidalgo,M., Saunders,A., Lawrence,M., Dietmann,S. *et al.* (2013) NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. *Nature*, **495**, 370–374.
64. Chen,T., Tsujimoto,N. and Li,E. (2004) The PWWP domain of Dnmt3a and Dnmt3b is required for directing DNA methylation to the major satellite repeats at pericentric heterochromatin. *Mol. Cell. Biol.*, **24**, 9048–9058.

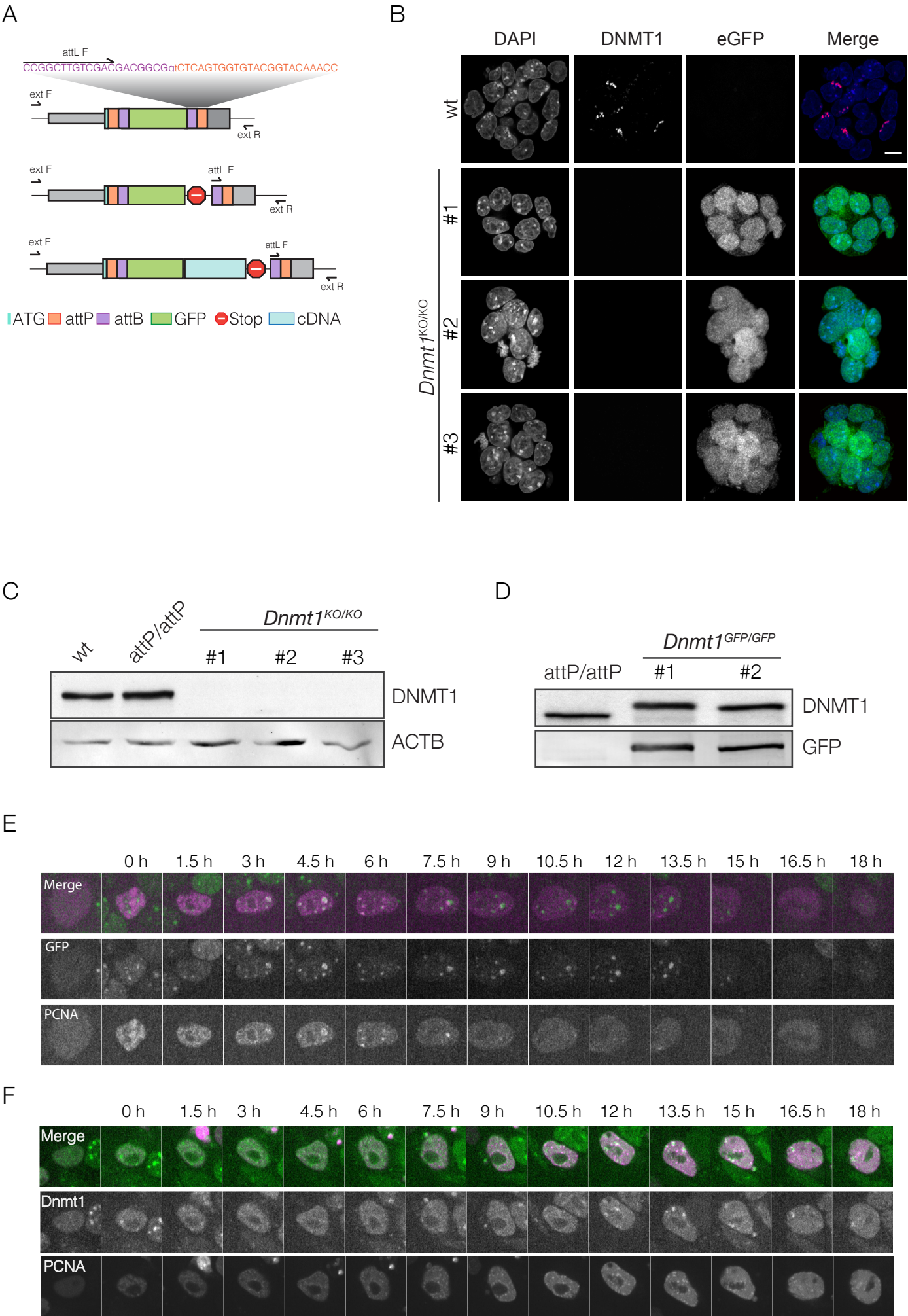
Supplementary Figure 1



Supplementary Figure 2



Supplementary Figure 3



Supplementary Figure 4

A

Expected	Seq_1	121	CAGCACATGGACAGCGGAGGAGGAGGC GAGAACCCTGTACTTT CAGGGAGGCCGAACCGGCT	180
Dnmt1 ^{GFP/GFP}	Seq_2	101	CAGCACATGGACAGCGGAGGAGGAGGC GAGAACCCTGTACTTT CAGGGAGGCCGAACCGGCT	160
attL Sequence				
Expected	Seq_1	181	TGTCGACGACGGCGGTCTCAGTGGTG TACGGTACAAACC ccagcgcgaacagctccagcc	240
Dnmt1 ^{GFP/GFP}	Seq_2	161	TGTCGACGACGGCGGTCTCAGTGGTG TACGGTACAAACC CCAGCGCAACAGCTCCAGCC	220
DNMT1 coding sequence				
Expected	Seq_1	241	cgagtgcctgcgcttgacctccccggcaggctcgctccccggaccatgtccgcaggcggttag	300
Dnmt1 ^{GFP/GFP}	Seq_2	221	CGAGTGCCTGCGCTTGCCCTCCCCGGCAGGCTCGCTCCCCGACCATGTCCGCAGGCGGTAG	280

B

Expected	Seq_1	121	CAGCACATGGACAGCGGAGGAGGAGCGAGAACCCTGTACTTTTCAGGGAGGCGGACCGGCT	180
Dnmt3b ^{GFP/GFP}	Seq_2	101	CAGCACATGGACAGCGGAGGAGGAGCGAGAACCCTGTACTTTTCAGGGAGGCGGACCGGCT	160
<hr/>				
Expected	Seq_1	181	TGTCGACGACGGCGGTCTCAGTGGTGTACGGTACAAACC	240
Dnmt3b ^{GFP/GFP}	Seq_2	161	TGTCGACGACGGCGGTCTCAGTGGTGTACGGTACAAACC	220
<hr/>				
attL Sequence				
<hr/>				
Expected	Seq_1	241	aatgaagaagaggggtgccagcgggtatgaggagtgcattatcgттаатgggaacttcagt	300
Dnmt3b ^{GFP/GFP}	Seq_2	221	AATGAAGAAGAGGGTGCCAGCGGGTATGAGGAGTGCATTATCGTTAATGGGAAC TTCAGT	280
<hr/>				
DNMT3B coding sequence				

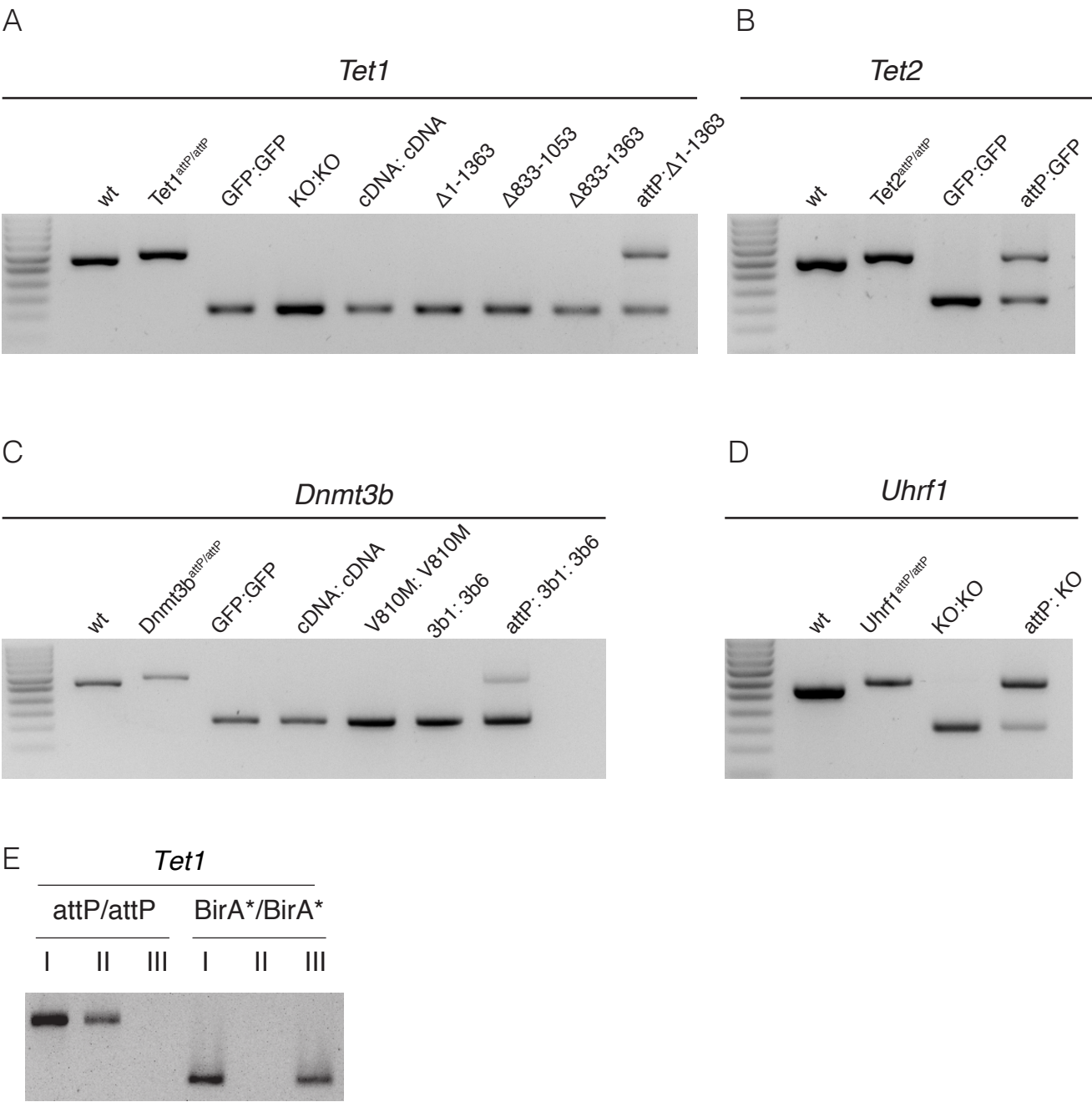
C

Expected	Seq_1	121	CAGCACATGGACAGCGGAGGAGGAGGC GAGAACCCTGTACTTTTCAGGGAGGCCGAACCGGCT	180
Tet1 ^{GFP/GFP}	Seq_2	101	CAGCACATGGACAGCGGAGGAGGAGGC GAGAACCCTGTACTTTTCAGGGAGGCCGAACCGGCT	160
Expected	Seq_1	181	TGTCGACGACGGCGGTCTCAGTGGTG TACGGTACAAACCcggtcccgcgccgcaaagcct	240
Tet1 ^{GFP/GFP}	Seq_2	161	TGTCGACGACGGCGGTCTCAGTGGTG TACGGTACAAACC CGGTCCCGCCCCGCAAAGCCT	220
attL Sequence				
Expected	Seq_1	241	tccaaatcagtcaaaaacaaagctacagaaaaagaagacatccagatgaagacgaagaca	300
Tet1 ^{GFP/GFP}	Seq_2	221	TCCAAATCAGTCAAAAACAAAGCTACAGAAAAAGAAAGACATCCAGATGAAGACGAAGACA	280
TET1 coding sequence				

D

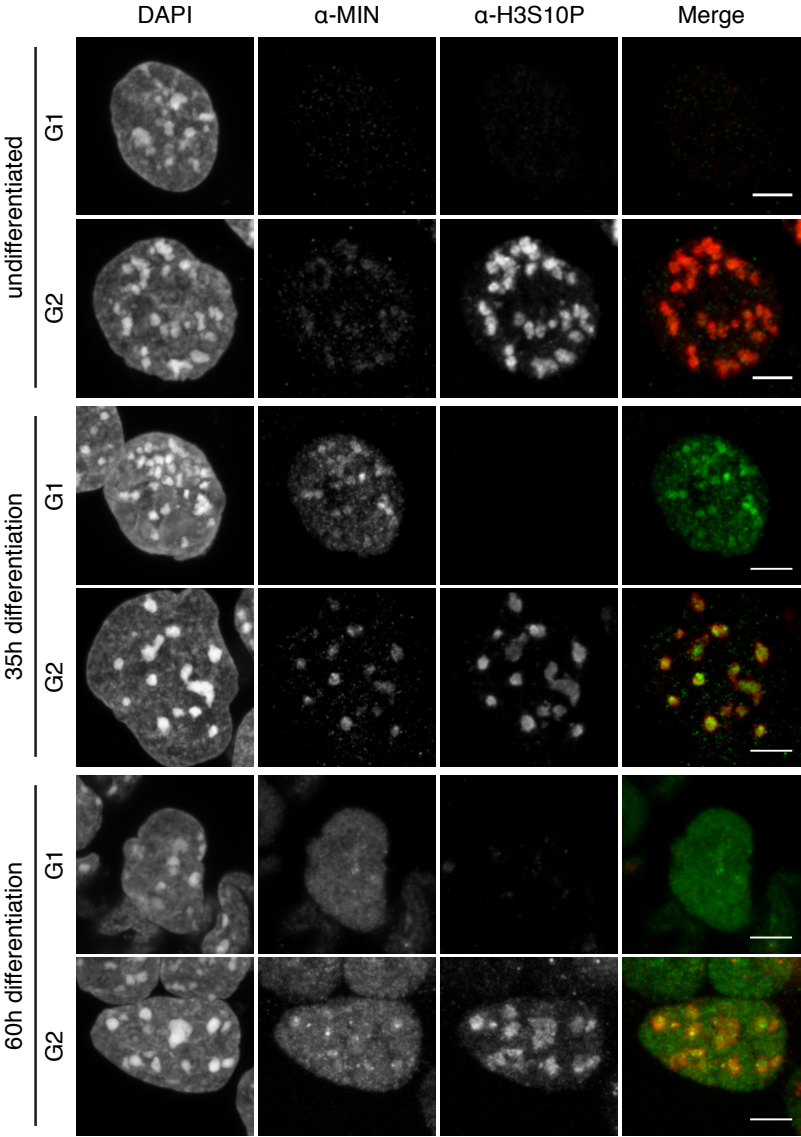
Expected	Seq_1	121	CAGCACATGGACAGCGGAGGAGGAGGC	GAGAACCTGTACTTT	CAGGGAGGC	CGGACCGGCT	180
Tet2 ^{GFP/GFP}	Seq_2	96	CAGCACATGGACAGCGGAGGAGGAGGC	GAGAACCTGTACTTT <td>CAGGGAGGC</td> <td>CGGACCGGCT</td> <td>155</td>	CAGGGAGGC	CGGACCGGCT	155
Expected	Seq_1	181	TGTCGACGACGGCGGTCTCAGTGGT	TACGGTACAAACC	gaacaggacagaaccacccat	240	
Tet2 ^{GFP/GFP}	Seq_2	156	TGTCGACGACGGCGGTCTCAGTGGT	TACGGTACAAACC	GAAGAGGACAGAACCACCCAT	215	
<i>attL</i> Sequence							
Expected	Seq_1	241	gctgagggcaccagactgagtcattcctgatagcaccaccttctcccatcagccataca	300			
Tet2 ^{GFP/GFP}	Seq_2	216	GCTGAGGGCACCAGACTGAGTCCATTCTCTGATAGCACCACCTTCTCCCATCAGCCATACA	275			
TET2 coding sequence							

Supplementary Figure 5

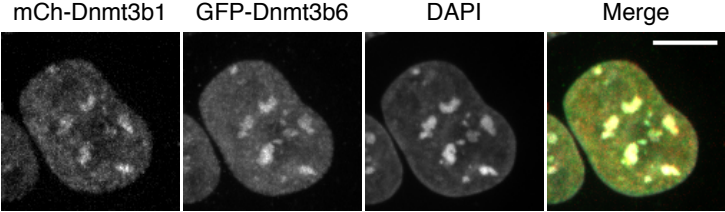


Supplementary figure 6

A



B



Supplemental Figure and Video Legends

Supplemental Figure S1. Characterization of MIN-tagged DNA methyltransferase cell lines.

(A) Western blot analysis of DNMT1 expression levels in the homozygous *Dnmt1*^{attP/attP} and wild type J1 cells. Beta-actin is used as a loading control. **(B)** Immunofluorescence stainings of Dnmt1 in wt and *Dnmt1*^{attP/attP} cells. Scale bar represent 5 μ m. **(C)** DNA methylation analysis of the major satellite repeats in *Dnmt1*^{attP/attP} and wild type cells. **(D)** Example of the screening PCRs, with and without HincII treatment, of clones found to be heterozygous and homozygous for the MIN-tag at the *Dnmt1* locus. Monoallelic and biallelic insertions of the MIN-tag can be distinguished by complete and incomplete digests, respectively. **(E)** Western blot analysis of DNMT3A expression levels in a heterozygous (#2) and homozygous (#1, #3-4) *Dnmt3a*^{attP/attP} cell lines compared to wild type cells. Beta-actin is used as a loading control. **(F)** DNA methylation analysis of major satellite repeats in *Dnmt3a*^{attP/attP} compared to wt cells. **(G)** Immunofluorescence stainings of DNMT3A together with the replication marker EdU in wt cells and the the homozygous *Dnmt3a*^{attP/attP} clone #1. Scale bar represents 10 μ m. **(H)** Immunofluorescence stainings of DNMT3B in *Dnmt3b*^{attP/attP} and wt cells after 35 hours of EpiLC differentiation. Scale bar represents 5 μ m. Error bar represent standard deviation (n=2).

Supplemental Figure S2. Characterization of MIN-tagged Tet1, Tet2 and Uhrf1 cell lines and C-terminal MIN-tag integration.

(A-C) Western blot analysis of TET1, TET2, and UHRF1 expression levels in the homozygous *Tet1*^{attP/attP}, *Tet2*^{attP/attP}, and N-terminal *Uhrf1*^{attP/attP} cell lines, respectively, compared to the wt J1 control. β -Actin (ACTB) was used as loading control. **(D)** Immunofluorescence stainings of TET1 in wt and *Tet1*^{attP/attP} cells. **(E)** Immunofluorescence stainings of 5-hydroxymethylcytosine (5-hmC) in wt and *Tet1*^{attP/attP} cells. **(F)** Immunofluorescence stainings of TET2 in wt and *Tet2*^{attP/attP} cells. **(G)**

Immunofluorescence stainings of UHRF1 in wt and *Uhrf1^{attP/attP}* cells. DAPI is used for DNA counterstaining; scale bars represent 15 μ m. **(H)** Schematic overview of CRISPR/Cas-assisted C-terminal integration of the MIN-tag. MIN-tag donors contain the attP site (depicted in orange) flanked by sequences (200-300 for PCR fragments or 76 for ssDNA oligos) homologous to 5' and 3' of the target gene stop codon (depicted in red). Restriction enzyme sites available for restriction fragment analysis based screening are shown above the attP sequence.

Supplemental Figure S3. Evaluating functionality of Bxb1 mediated recombination in *Dnmt1^{attP/attP}* cells.

(A) Schematic outline of the multiplex PCR strategy to identify positive recombination events and their zygosity. **(B)** Immunofluorescence stainings of DNMT1 and GFP in wt cells and three *Dnmt1^{KO/KO}* clones. Diffuse GFP indicates a successful integration of the KO cassette into the locus. **(C)** Western blot analysis of DNMT1 expression levels in three *Dnmt1^{KO/KO}* clonal cell lines generated by Bxb1-mediated insertion of a knock-out cassette, compared to wt and *Dnmt1^{attP/attP}* cells. **(D)** Western blot analysis of DNMT1 and GFP expression in *Dnmt1^{attP/attP}* cells and two homozygous GFP-knock in cell lines (*Dnmt1^{GFP/GFP}* #1-2) generated by Bxb1 mediated insertion. **(E-F)** Live cell imaging of *Dnmt1^{GFP/GFP}* and *Dnmt1^{cDNA/cDNA}* cells transiently expressing RFP-labeled PCNA, a DNA replication marker, during cell-cycle progression. Scale bars represent 5 μ m

Supplemental Figure S4. Alignments of the expected sequence flanking the attL site after recombination

Alignments of the expected sequence flanking the attL site after recombination of the attB-GFP KI at the *Dnmt1*, *Dnmt3b*, *Tet1*, and *Tet2* locus (A-D) with the sequencing results from the *Dnmt1^{GFP/GFP}*, *Dnmt3b^{GFP/GFP}*, *Tet1^{GFP/GFP}*, and *Tet2^{GFP/GFP}* cell lines.

Supplemental Figure S5. Demonstration of Bxb1 mediated recombination in multiple MIN-tagged genes.

(A-D) Gel electrophoresis of the multiplex PCR (using the attL primer and locus specific

external primers, see also Table S1) performed on cell lines generated by Bxb1-mediated integration of various MIN-tag toolbox components (Table S5) into the loci of: (A) Tet1, (B) Tet2, (C) Dnmt3b, and (D) Uhrf1. Equal mixtures of genomic DNA from non-recombined cell lines and recombined cell lines are used to control for possible amplification biases arising from the use of different locus specific external primers. **(E)** PCR to confirm insertion of the BirA* cassette into the Tet1 genomic locus. I: multiplex PCR, II: wt specific PCR, III: attL (recombination) specific PCR

Supplemental Figure S6. Cell cycle analysis of DNMT3b localization during differentiation.

(A) Immunofluorescence stainings of MIN-tagged DNMT3B and Histone 3 Serine 10 phosphorylation (H3S10P), a marker of G2/M phase (Ref Hendzel:1997wo) during differentiation of naive pluripotent *Dnmt3b*^{attP/attP} stem cells into epiblast-like cells. Cells were fixed directly after (0 h)35 h, or 60 h after induction of differentiation. The H3S10P mark was used to determine if cells were in G2 or G1 phase in order to assess whether changes in DNMT3B localization during differentiation are cell-cycle dependent. Scale bar represents 5 μ m. **(B)** Fluorescence microscopy images of *Dnmt3b*^{mCh-3b1/GFP-3b6} cells fixed after 35 h of differentiation. Both DNMT3B isoforms (GFP-DNMT3B1 in green and mCh-DNMT3B6 in red) localize at chromocenters (visible as bright DAPI spots). Scale bar represents 5 μ m

Supplemental Video 1. Live cell imaging of *Dnmt3b*^{GFP/GFP} cells during differentiation.

Long-term (60 h), live cell imaging tracking the transition of *Dnmt3b*^{GFP/GFP} cells from the naive pluripotency ground state into the primed, epiblast-like state. Images were acquired once per hour and entailing at least 10 μ m z-stacks. The left panel depicts the projection of GFP signal, while the right panel shows that projection superimposed onto the acquired brightfield images.

Supplemental Tables (S1-S5)

Table S1: CRISPR/Cas9-mediated MIN-tag insertion efficiencies

Gene	Position	MIN-tag Donor	Heterozygotes	Homozygotes	TOTAL
Dnmt1	N-terminal	PCR Product	1/67 (1.5%)	1/67 (1.5%)	2/67 (2.9%)
Dnmt3a	N-terminal	PCR Product	0/86 (0%)	3/86 (3.5%)	3/86 (3.5%)
Dnmt3b	N-terminal	ssDNA oligo	0/65(0%)	1/65(1.5%)	1/65 (1.5%)
Uhrf1	N-terminal	PCR Product	0/6 (0%)	1/6(16.7%)	1/6 (16.7%)
Uhrf1	C-terminal	ssDNA oligo	2/36 (5.5%)	2/36 (5.6 %)	4/36 (11.1%)
Tet1	N-terminal	PCR Product	0/70(0%)	1/70 (1.4%)	1/70 (1.4%)
Tet2	N-terminal	PCR Product	1/24 (4.2%)	2/24 (8.3%)	3/24 (12.5%)
Tet3	N-terminal	PCR Product	0/38 (0%)	2/38 (5.3%)	2/38 (5.3%)

Table S2: Oligonucleotide sequences used for CRISPR/Cas assisted targeting and screening

Name	Sequence
Dnmt1	
gRNA_F	TGTTTCGCGCTGGCATCTTGCGTTTTAGAGCTAGAAATAGCAAG
gRNA_R	GCAAGATGCCAGCGCGAACACGGTGTTTCGTCCTTTCCAC
surrogate_F	CTAGCTGTTTCGCGCTGGCATCTTGAGGGGATTCC
surrogate_R	CCGGAGGAATCCCCTGCAAGATGCCAGCGCGAACAG
internal_R	CACTATAGCCAGGAGGTGTGGG
internal_F	TGTACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCCATCTTGCAGGTTGCA GACGACAG
external_R	GTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCCAGCGCGAACAGCTCC AGC
external_F	GCGCGACAGGAAGCACAGCC
screening_F	GTCGCAGCACGGACGAG
Uhrf1 (N)	
gRNA_F	CATCGGCATCATGTGGATCCGTTTTAGAGCTAGAAATAGCAAG
gRNA_R	GGATCCACATGATGCCGATGCGGTGTTTCGTCCTTTCCAC
surrogate_F	CTAGCCATCGGCATCATGTGGATCCAGGGGATTCCT
surrogate_R	GGCCAGGAATCCCCTGGATCCACATGATGCCGATGG
internal_R	CATCGGCATCATGTGGATCCGTTTTAGAGCTAGAAATAGCAAG
internal_F	GGATCCACATGATGCCGATGCGGTGTTTCGTCCTTTCCAC
external_R	ACCACCGCGGTCTCAGTGGTGTACGGTACAAACCTGGATCCAGGTTCGAACTATG
external_F	CTATTGCTTGGTGGCTTTGAG
screening_F	GGCAATTCACATTCAAGTGTCCC
Uhrf1 (C)	
gRNA_F	TGCCTGGGTCTCAGCATCACGTTTTAGAGCTAGAAATAGCAAG
gRNA_R	GTGATGCTGAGACCCAGGCACGGTGTTTCGTCCTTTCCAC
surrogate_F	CTAGCTGCCTGGGTCTCAGCATCACCGGGGATTCCT
surrogate_R	CCGGAGGAATCCCCGGTGATGCTGAGACCCAGGCAG
ssDNA oligo	CAGCTCCCCAACCCGGGTGAACCAGCCCTTGCAGACCATTCTCAACCAGCTCTTCCCTGG CTATGGCAGCGGCCGGGTTTGTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACA AACCTGATGCTGAGACCCAGGCAGAGGGCTCATGGTTCCAATTTCATAGTGTGTTAGCT TGAAGGTGTTGTCCTTCACG
external_R	TTTCTAGGCAGCTGGTGTGG
external_F	TGTACGTGAGAGGACGGAGT
screening_F	TGTTGCCAGGAGCTACCAAG
Dnmt3a	
gRNA_F	GGGCCGCTGGAGGGCATTGCGTTTTAGAGCTAGAAATAGCAAG
gRNA_R	GCAATGCCCTCCAGCGGCCCGGTGTTTCGTCCTTTCCAC
surrogate_F	CTAGCGGGCCGCTGGAGGGCATTGCTGGGGATTCCT
surrogate_R	CCGGAGGAATCCCCAGCAATGCCCTCCAGCGGCCCG
internal_R	CTTCTCTTCCCCACAGGCAG
internal_F	ACCACTGAGACCGCGGTGGTTGACCAGACAAACCCATTGCTGGGCAGTAGGCG
external_R	ACCACCGCGGTCTCAGTGGTGTACGGTACAAACCCCTCCAGCGGCCCG
external_F	GTTCCCAGCCAAGCACCTAT
screening_F	ATGGTCCTGCAACCAGAGTG
Dnmt3b	
gRNA_F	TTCCCCACAGGAAACAATGAGTTTTAGAGCTAGAAATAGCAAG
gRNA_R	TCATTGTTTCCTGTGGGAACGGTGTTTCGTCCTTTCCAC

surrogate_F	CTAGCTTCCCCACAGGAAACAATGAAGGGGATTCCT
surrogate_R	CCGGAGGAATCCCCTTCATTGTTTCTGTGGGGAAG
ssDNA oligo	GAACTGGTGGTGTAACCTTGCAGTGTGCCCTGTCTGCCTCTTACATATCCTGATCTTTC CCCACAGGAAACAATGGGTTTGTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACA AACCAAGGGAGACAGCAGACATCTGAATGAAGAAGAGGGTGCCAGCGGGTATGAGGAGTG CATTATCGTTAATGGGAACT
external_R	ACCACCGCGGTCTCAGTGGTGTACGGTACAAACCGGAGACAGCAGACATCTGAATG
external_F	ATCTGTTCATGGAACCTGCCG
screening_F	GAGCTGGCCAATTGCAGAAC
Tet1	
gRNA_F	AGACATGGCTGCAGAGTAAGCGGTGTTTCGTCCTTTCCAC
gRNA_R	CTTACTCTGCAGCCATGTCTAGCTTTCTTGTACAAAGTTGGCAT
surrogate_F	CTAGCCTTACTCTGCAGCCATGTCTCGGGGATCCCT
surrogate_R	CCGGAGGGATCCCCGAGACATGGCTGCAGAGTAAGG
internal_R	ACTCAGTCTCCCAAATGCTGG
internal_F	ACCACTGAGACCGCGGTGGTTGACCAGACAAACCAGACATGGCTGCAGAGTAAGTAAAG
external_R	ACCACCGCGGTCTCAGTGGTGTACGGTACAAACCGGTCCCGCCCCGCAAAG
external_F	TCGGGGTTTTGTCTTCCGTT
screening_F	GGGCAATGTTGTGACTCATGC
Tet2	
gRNA_F	CGAAGCAAGCCTGATGGAACGTTTTAGAGCTAGAAATAGCAAG
gRNA_R	GTTCCATCAGGCTTGCTTCGCGGTGTTTCGTCCTTTCCAC
surrogate_F	CTAGCCGAAGCAAGCCTGATGGAACAGGGGATTCCT
surrogate_R	CCGGAGGAATCCCCTGTTCCATCAGGCTTGCTTCGG
internal_R	ACCACTGAGACCGCGGTGGTTGACCAGACAAACCCATCAGGCTTGCTTCGGGG
internal_F	ACCACCGCGGTCTCAGTGGTGTACGGTACAAACCGAACAGGACAGAACCACCCAT
external_R	TGGTTCACTGACTGTGCGTT
external_F	CCAGGATCACACAGGAAGCA
screening_F	GGATGGAGCCCAGAGAGAGA
Tet3	
gRNA_F	GTTCCAGGTCAGATGGACTCGTTTTAGAGCTAGAAATAGCAAG
gRNA_R	GAGTCCATCTGACCTGGAACCGGTGTTTCGTCCTTTCCAC
surrogate_F	CTAGCGTTCCAGGTCAGATGGACTCAGGGGATTCCT
surrogate_R	CCGGAGGAATCCCCTGAGTCCATCTGACCTGGAACG
internal_R	ACCACTGAGACCGCGGTGGTTGACCAGACAAACCCATCTGACCTGGAACAGGTC
internal_F	ACCACCGCGGTCTCAGTGGTGTACGGTACAAACCGACTCAGGGCCAGTGTACC
external_R	CAGTCGGGCTTCTGGTCTAC
external_F	GATCTGAGCTCTCACAGGGC
screening_F	AGTAGACAGGGCCTTGGGAT
attL_F	CCGGCTTGTCGACGACG

Table S3: Bxb1-mediated recombination efficiencies

Gene	Integration Construct	Heterozygotes	Homozygotes	TOTAL
Dnmt1	attB-GFP	N/A	13/31 (41.9%)	13/31 (41.9%)
Dnmt3b	attB-GFP	0/3 (0%)	1/3 (33.3%)	1/3 (33.3%)
Tet1	attB-GFP	14/45 (31.1%)	13/45 (28.9%)	27/45 (60%)
Tet2	attB-GFP	28/81 (34.6%)	15/81 (18.5%)	43/81 (53%)
Dnmt1	attB-GFP-STOP-Poly(A)	2/23 (8.7%)	13/23 (56.5%)	15/23 (65.2%)
Uhrf1	attB-GFP-STOP-Poly(A)	5/32 (15.6%)	14/32 (43.8%)	19/32 (59.4%)
Dnmt1	attB-GFP-cDNA-STOP-Poly(A)	1/15 (6.6%)	9/15 (60%)	10/15 (66.6%)
Dnmt3b	attB-GFP-cDNA-STOP-Poly(A)	28/84 (33.3%)	26/84 (31%)	54/84 (64.3%)
Tet1	attB-GFP-cDNA-STOP-Poly(A)	12/58 (20.7%)	7/58 (12.1%)	19/58 (32.8%)
Dnmt3b	attB-GFP/mCh-cDNA-STOP-Poly(A) PuroR/neoR	29/102 (28.4%)	64/102 (62.7%)	93/102 (91.2%)

Table S4: Evaluation of FRAP protein kinetics

	GFP-DNMT3B	mCh-DNMT3B1	GFP-DNMT3B6
Mobile fraction [A]	87	81	100
Diffusion coef. [$\mu\text{m}^2/\text{s}$]	4.2E-03	1.2E-03	4.1E-02
Half-time recovery [s]	42.2	94.8	5.1

Table S5: The MIN-tag toolbox

Name	Fluorescent protein	Application
Universal constructs		
attB-GFP	GFP	GFP KI
attB-mCh	mCherry	mCherry KI
attB-GFP-T2A-BirA*	GFP	Protein interaction
attB-GFP-Poly(A)	GFP	KO
attB-mCh-Poly(A)	mCherry	KO
attB-GFP-Poly(A)-NeoR	GFP	KO /w selection
attB-GFP-Poly(A)-PuroR	GFP	KO /w selection
attB-mCh-Poly(A)-NeoR	mCherry	KO /w selection
attB-mCh-Poly(A)-PuroR	mCherry	KO /w selection
Gene specific cDNA KI constructs		
attB-GFP-Dnmt1-Poly(A)	GFP	cDNA KI
attB-GFP-Dnmt3b1-Poly(A)	GFP	cDNA KI
attB-GFP-Dnmt3b6-Poly(A)	GFP	cDNA KI
attB_eGFP_Dnmt3b_C656A_Poly(A)	GFP	cDNA KI
attB_eGFP_Dnmt3b_D809G_Poly(A)	GFP	cDNA KI
attB_eGFP_Dnmt3b_dX_Poly(A)	GFP	cDNA KI
attB_eGFP_Dnmt3b_G655S_Poly(A)	GFP	cDNA KI
attB_eGFP_Dnmt3b_L656T_Poly(A)	GFP	cDNA KI
attB_eGFP_Dnmt3b_V718G_Poly(A)	GFP	cDNA KI
attB_eGFP_Dnmt3b_V810M_Poly(A)	GFP	cDNA KI
attB_eGFP_Dnmt3b6_Poly(A)	GFP	cDNA KI
attB_eGFP_Dnmt3b1_dPWWP_Poly(A)	GFP	cDNA KI
attB_eGFP_Dnmt3b1_dPHD_Poly(A)	GFP	cDNA KI
attB_mCh_Dnmt3b_C656A_Poly(A)	mCherry	cDNA KI
attB_mCh_Dnmt3b_D809G_Poly(A)	mCherry	cDNA KI
attB_mCh_Dnmt3b_dX_Poly(A)	mCherry	cDNA KI
attB_mCh_Dnmt3b_G655S_Poly(A)	mCherry	cDNA KI
attB_mCh_Dnmt3b_L656T_Poly(A)	mCherry	cDNA KI
attB_mCh_Dnmt3b_V718G_Poly(A)	mCherry	cDNA KI
attB_mCh_Dnmt3b_V810M_Poly(A)	mCherry	cDNA KI
attB_mCh_Dnmt3b6_Poly(A)	mCherry	cDNA KI
attB-GFP-Dnmt3b1-Poly(A) -NeoR	GFP	cDNA KI /w selection
attB-GFP-Dnmt3b6-Poly(A) -NeoR	GFP	cDNA KI /w selection
attB_eGFP_Dnmt3b_C656A_Poly(A)-NeoR	GFP	cDNA KI /w selection
attB_eGFP_Dnmt3b_D809G_Poly(A)-NeoR	GFP	cDNA KI /w selection
attB_eGFP_Dnmt3b_dX_Poly(A)-NeoR	GFP	cDNA KI /w selection
attB_eGFP_Dnmt3b_G655S_Poly(A)-NeoR	GFP	cDNA KI /w selection
attB_eGFP_Dnmt3b_L656T_Poly(A)-NeoR	GFP	cDNA KI /w selection
attB_eGFP_Dnmt3b_V718G_Poly(A)-NeoR	GFP	cDNA KI /w selection

attB_eGFP_Dnmt3b_V810M_Poly(A)-NeoR	GFP	cDNA KI /w selection
attB_eGFP_Dnmt3b6_Poly(A)-NeoR	GFP	cDNA KI /w selection
attB- mCh -Dnmt3b1-Poly(A) -NeoR	mCherry	cDNA KI /w selection
attB- mCh -Dnmt3b6-Poly(A) -NeoR	mCherry	cDNA KI /w selection
attB_mCh_Dnmt3b_C656A_Poly(A)-NeoR	mCherry	cDNA KI /w selection
attB_mCh_Dnmt3b_D809G_Poly(A)-NeoR	mCherry	cDNA KI /w selection
attB_mCh_Dnmt3b_dX_Poly(A)-NeoR	mCherry	cDNA KI /w selection
attB_mCh_Dnmt3b_G655S_Poly(A)-NeoR	mCherry	cDNA KI /w selection
attB_mCh_Dnmt3b_L656T_Poly(A)-NeoR	mCherry	cDNA KI /w selection
attB_mCh_Dnmt3b_V718G_Poly(A)-NeoR	mCherry	cDNA KI /w selection
attB_mCh_Dnmt3b_V810M_Poly(A)-NeoR	mCherry	cDNA KI /w selection
attB_mCh_Dnmt3b6_Poly(A)-PuroR	mCherry	cDNA KI /w selection
attB- mCh -Dnmt3b1-Poly(A)-PuroR	mCherry	cDNA KI /w selection
attB- mCh -Dnmt3b6-Poly(A)-PuroR	mCherry	cDNA KI /w selection
attB_mCh_Dnmt3b_C656A_Poly(A)-PuroR	mCherry	cDNA KI /w selection
attB_mCh_Dnmt3b_D809G_Poly(A)-PuroR	mCherry	cDNA KI /w selection
attB_mCh_Dnmt3b_dX_Poly(A)- PuroR	mCherry	cDNA KI /w selection
attB_mCh_Dnmt3b_G655S_Poly(A)-PuroR	mCherry	cDNA KI /w selection
attB_mCh_Dnmt3b_L656T_Poly(A)-PuroR	mCherry	cDNA KI /w selection
attB_mCh_Dnmt3b_V718G_Poly(A)-PuroR	mCherry	cDNA KI /w selection
attB_mCh_Dnmt3b_V810M_Poly(A)-PuroR	mCherry	cDNA KI /w selection
attB_mCh_Dnmt3b6_Poly(A)-PuroR	mCherry	cDNA KI /w selection
attB-GFP-Tet1-Poly(A)	GFP	cDNA KI
attB-GFP-Tet1d1-389-Poly(A)	GFP	cDNA KI
attB-GFP-Tet1d390-565-Poly(A)	GFP	cDNA KI
attB-GFP-Tet1d566-833-Poly(A)	GFP	cDNA KI
attB-GFP-Tet1d834-1053-Poly(A)	GFP	cDNA KI
attB-GFP-Tet1d1054-1363-Poly(A)	GFP	cDNA KI
attB-GFP-Tet1d1-833-Poly(A)	GFP	cDNA KI
attB-GFP-Tet1d834-1363-Poly(A)	GFP	cDNA KI
attB-GFP-Tet2-Poly(A)	GFP	cDNA KI
attB-GFP-Tet2d1-225-Poly(A)	GFP	cDNA KI
attB-GFP-Tet2d226-398-Poly(A)	GFP	cDNA KI
attB-GFP-Tet2d399-650-Poly(A)	GFP	cDNA KI
attB-GFP-Tet2d651-848-Poly(A)	GFP	cDNA KI
attB-GFP-Tet2d849-1038-Poly(A)	GFP	cDNA KI
attB-GFP-Tet2d1-650-Poly(A)	GFP	cDNA KI
attB-GFP-Tet2d651-1038-Poly(A)	GFP	cDNA KI
attB-GFP-Uhrf1-Poly(A)	GFP	cDNA KI
attB-GFP-Uhrf1dSRA-Poly(A)	GFP	cDNA KI

Supplemental Table Legends

Table S1: CRISPR/Cas9-mediated MIN-tag insertion efficiencies

For MIN-tag Insertion, J1 mESCs transfected with the appropriate MIN-tag donor oligonucleotides or PCR products along with the Cas9, gRNA, and CRISPR surrogate reporter vector were single cell sorted after enriching for cells with CRISPR/Cas activity. The number of clones with either a monoallelic or biallelic insertion of the MIN-Tag is shown in relation to the number of clones screened.

Table S2: Oligonucleotide sequences used for CRISPR/Cas assisted targeting and screening

DNA oligonucleotides used for the generation of target specific gRNA expression vectors, surrogate reporters, and homology donors for MIN-tag integration.

Table S3: Bxb1-mediated recombination efficiencies

For Bxb1-mediated recombination, J1 mESCs transfected with NLS-Bxb1, the Bxb1 surrogate reporter, and the respective attB-site containing integration construct were single-cell sorted after enrichment for cells with Bxb1 activity. The number of clones with either a monoallelic or biallelic integration of the listed construct is shown in relation to the total number of clones screened.

Table S4: Evaluation of FRAP protein kinetics

Evaluation of FRAP kinetics (w/o 5-azadC treatment) performed in Dnmt3bGFP/GFP and Dnmt3bmCh-3b1/GFP-3b6 cells

Table S5: The MIN-tag toolbox

Vectors generated for Bxb1 mediated recombination into MIN-tagged cell lines. KO: knockout, KI: knockin

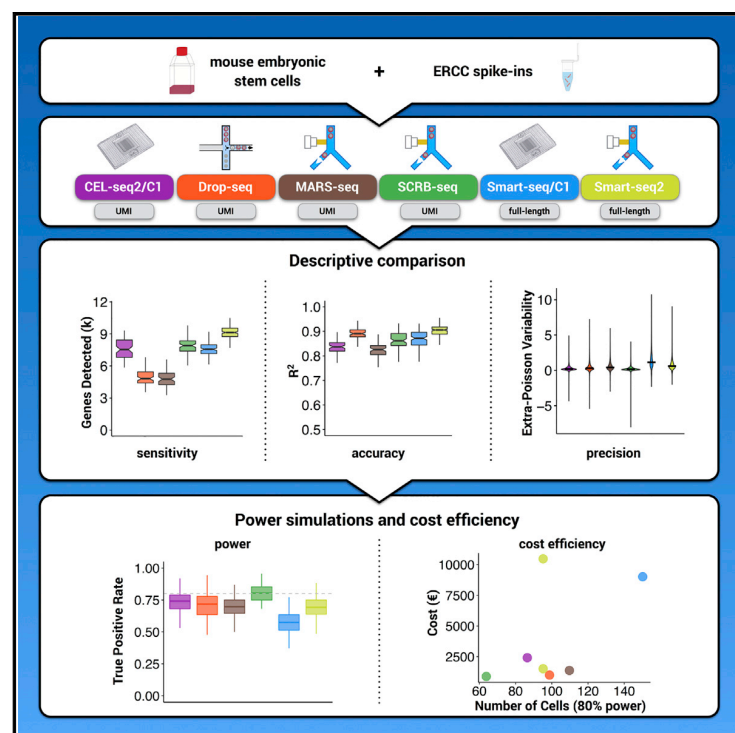
Results

2.4 Comparative analysis of single-cell RNA sequencing methods

Molecular Cell

Comparative Analysis of Single-Cell RNA Sequencing Methods

Graphical Abstract



Authors

Christoph Ziegenhain, Beate Vieth, Swati Parekh, ..., Holger Heyn, Ines Hellmann, Wolfgang Enard

Correspondence

enard@bio.lmu.de

In Brief

Ziegenhain et al. generated data from mouse ESCs to systematically evaluate six prominent scRNA-seq methods. They used power simulations to compare cost efficiencies, allowing for informed choice among existing protocols and providing a framework for future comparisons.

Highlights

- The study represents the most comprehensive comparison of scRNA-seq protocols
- Power simulations quantify the effect of sensitivity and precision on cost efficiency
- The study offers an informed choice among six prominent scRNA-seq methods
- The study provides a framework for benchmarking future protocol improvements



Ziegenhain et al., 2017, Molecular Cell 65, 631–643
February 16, 2017 © 2017 Elsevier Inc.
<http://dx.doi.org/10.1016/j.molcel.2017.01.023>

CellPress

Comparative Analysis of Single-Cell RNA Sequencing Methods

Christoph Ziegenhain,¹ Beate Vieth,¹ Swati Parekh,¹ Björn Reinius,^{2,3} Amy Guillaumet-Adkins,^{4,5} Martha Smets,⁶ Heinrich Leonhardt,⁶ Holger Heyn,^{4,5} Ines Hellmann,¹ and Wolfgang Enard^{1,7,*}

¹Anthropology & Human Genomics, Department of Biology II, Ludwig-Maximilians University, Großhaderner Straße 2, 82152 Martinsried, Germany

²Ludwig Institute for Cancer Research, Box 240, 171 77 Stockholm, Sweden

³Department of Cell and Molecular Biology, Karolinska Institutet, 171 77 Stockholm, Sweden

⁴CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), 08028 Barcelona, Spain

⁵Universitat Pompeu Fabra (UPF), 08002 Barcelona, Spain

⁶Department of Biology II and Center for Integrated Protein Science Munich (CIPSM), Ludwig-Maximilians University, Großhaderner Straße 2, 82152 Martinsried, Germany

⁷Lead Contact

*Correspondence: enard@bio.lmu.de

<http://dx.doi.org/10.1016/j.molcel.2017.01.023>

SUMMARY

Single-cell RNA sequencing (scRNA-seq) offers new possibilities to address biological and medical questions. However, systematic comparisons of the performance of diverse scRNA-seq protocols are lacking. We generated data from 583 mouse embryonic stem cells to evaluate six prominent scRNA-seq methods: CEL-seq2, Drop-seq, MARS-seq, SCRB-seq, Smart-seq, and Smart-seq2. While Smart-seq2 detected the most genes per cell and across cells, CEL-seq2, Drop-seq, MARS-seq, and SCRB-seq quantified mRNA levels with less amplification noise due to the use of unique molecular identifiers (UMIs). Power simulations at different sequencing depths showed that Drop-seq is more cost-efficient for transcriptome quantification of large numbers of cells, while MARS-seq, SCRB-seq, and Smart-seq2 are more efficient when analyzing fewer cells. Our quantitative comparison offers the basis for an informed choice among six prominent scRNA-seq methods, and it provides a framework for benchmarking further improvements of scRNA-seq protocols.

INTRODUCTION

Genome-wide quantification of mRNA transcripts is highly informative for characterizing cellular states and molecular circuitries (ENCODE Project Consortium, 2012). Ideally, such data are collected with high spatial resolution, and single-cell RNA sequencing (scRNA-seq) now allows for transcriptome-wide analyses of individual cells, revealing exciting biological and medical insights (Kolodziejczyk et al., 2015a; Wagner et al., 2016). scRNA-seq requires the isolation and lysis of single cells, the conversion of their RNA into cDNA, and the amplification of cDNA to generate high-throughput sequencing libraries. As the

amount of starting material is so small, this process results in substantial technical variation (Kolodziejczyk et al., 2015a; Wagner et al., 2016).

One type of technical variable is the sensitivity of a scRNA-seq method (i.e., the probability to capture and convert a particular mRNA transcript present in a single cell into a cDNA molecule present in the library). Another variable of interest is the accuracy (i.e., how well the read quantification corresponds to the actual concentration of mRNAs), and a third type is the precision with which this amplification occurs (i.e., the technical variation of the quantification). The combination of sensitivity, precision, and number of cells analyzed determines the power to detect relative differences in expression levels. Finally, the monetary cost to reach a desired level of power is of high practical relevance. To make a well-informed choice among available scRNA-seq methods, it is important to quantify these parameters comparably. Some strengths and weaknesses of different methods are already known. For example, it has previously been shown that scRNA-seq conducted in the small volumes available in the automated microfluidic platform from Fluidigm (C1 platform) outperforms CEL-seq2, Smart-seq, or other commercially available kits in microliter volumes (Hashimshony et al., 2016; Wu et al., 2014). Furthermore, the Smart-seq protocol has been optimized for sensitivity, more even full-length coverage, accuracy, and cost (Picelli et al., 2013), and this improved Smart-seq2 protocol (Picelli et al., 2014b) has also become widely used (Gokce et al., 2016; Reinius et al., 2016; Tirosh et al., 2016).

Other protocols have sacrificed full-length coverage in order to sequence part of the primer used for cDNA generation. This enables early barcoding of libraries (i.e., the incorporation of cell-specific barcodes), allowing for multiplexing the cDNA amplification and thereby increasing the throughput of scRNA-seq library generation by one to three orders of magnitude (Hashimshony et al., 2012; Jaitin et al., 2014; Klein et al., 2015; Macosko et al., 2015; Soumillon et al., 2014). Additionally, this approach allows the incorporation of unique molecular identifiers (UMIs), random nucleotide sequences that tag individual

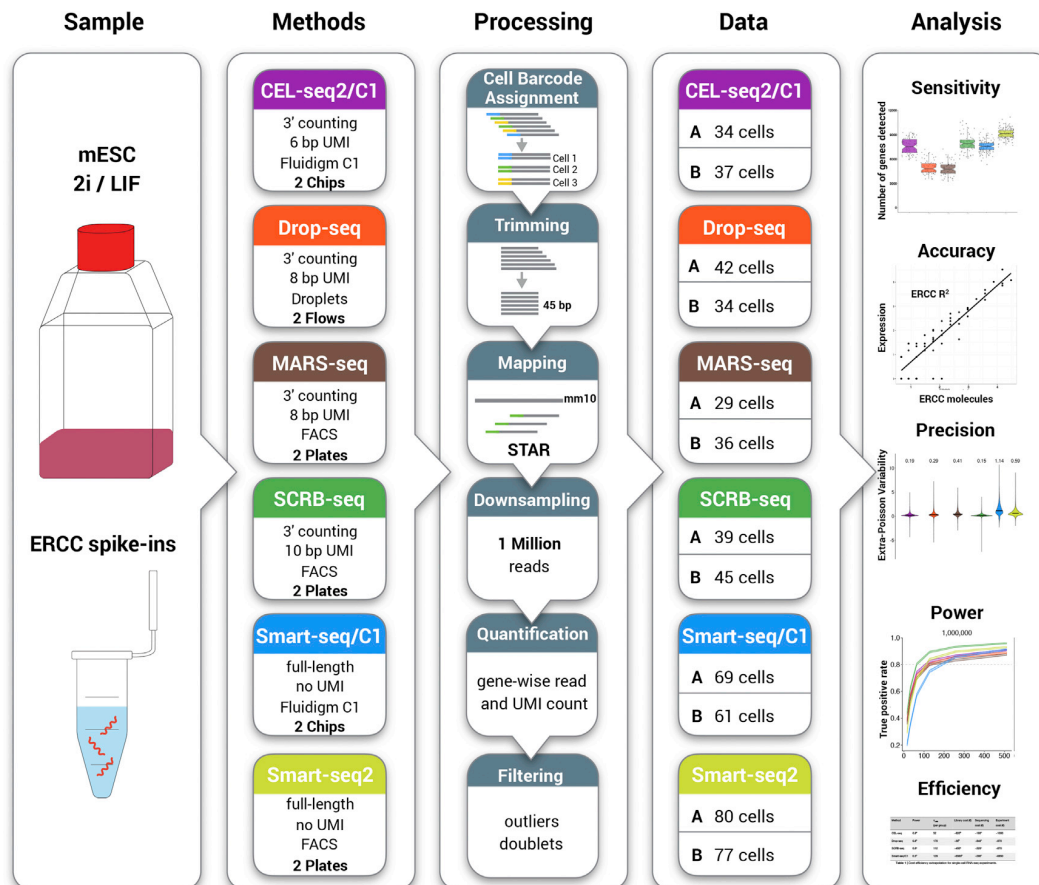


Figure 1. Schematic Overview of the Experimental and Computational Workflow

Mouse embryonic stem cells (mESCs) cultured in 2i/LIF and ERCC spike-in RNAs were used to generate single-cell RNA-seq data with six different library preparation methods (CEL-seq2/C1, Drop-seq, MARS-seq, SCR-seq, Smart-seq/C1, and Smart-seq2). The methods differ in the usage of unique molecular identifier (UMI) sequences, which allow the discrimination between reads derived from original mRNA molecules and duplicates generated during cDNA amplification. Data processing was identical across methods, and the given cell numbers per method and replicate were used to compare sensitivity, accuracy, precision, power, and cost efficiency. The six scRNA-seq methods are denoted by color throughout the figures of this study as follows: purple, CEL-seq2/C1; orange, Drop-seq; brown, MARS-seq; green, SCR-seq; blue, Smart-seq; and yellow, Smart-seq2. See also Figures S1 and S2.

mRNA molecules and, hence, allow for the distinction between original molecules and amplification duplicates that derive from the cDNA or library amplification (Kivioja et al., 2011). Utilization of UMI information improves quantification of mRNA molecules (Grün et al., 2014; Islam et al., 2014), and it has been implemented in several scRNA-seq protocols, such as STRT (Islam et al., 2014), CEL-seq (Grün et al., 2014; Hashimshony et al., 2016), CEL-seq2 (Hashimshony et al., 2016), Drop-seq (Macosko et al., 2015), inDrop (Klein et al., 2015), MARS-seq (Jaitin et al., 2014), and SCR-seq (Soumillon et al., 2014).

However, a thorough and systematic comparison of relevant parameters across scRNA-seq methods is still lacking. To address this issue, we generated 583 scRNA-seq libraries from mouse embryonic stem cells (mESCs), using six different methods in two replicates, and we compared their sensitivity, accuracy, precision, power, and efficiency (Figure 1).

RESULTS

Generation of scRNA-Seq Libraries

Variation in gene expression as observed among single cells is caused by biological and technical variation (Kolodziejczyk et al., 2015a; Wagner et al., 2016). We used mESCs cultured under two inhibitor/leukemia inhibitory factor (2i/LIF) conditions to obtain a relatively homogeneous cell population (Grün et al., 2014; Kolodziejczyk et al., 2015b), so that biological variation was similar among experiments and, hence, we mainly compared technical variation. In addition, we spiked in 92 poly-adenylated synthetic RNA transcripts of known concentration designed by the External RNA Control Consortium (ERCCs) (Jiang et al., 2011). For all six tested scRNA-seq methods (Figure 2), we generated libraries in two independent replicates.

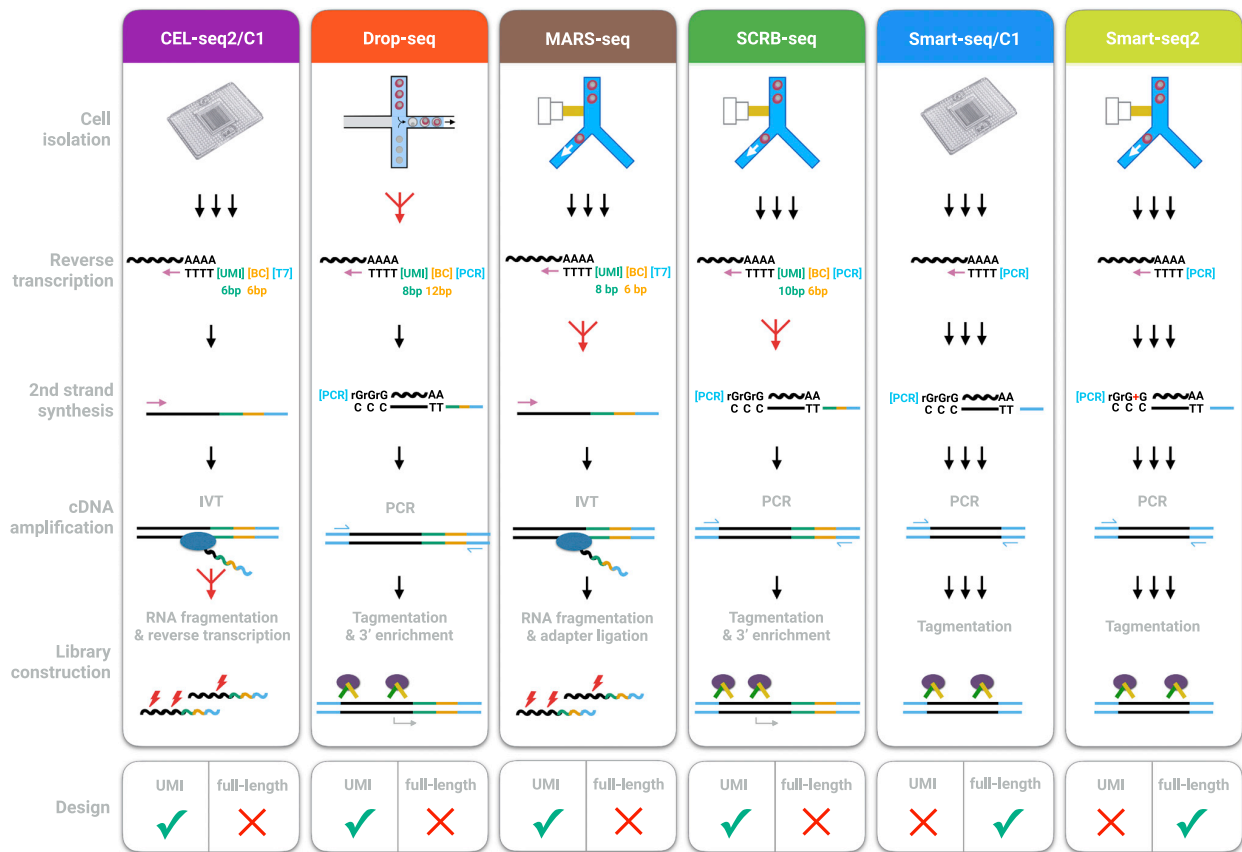


Figure 2. Schematic Overview of Library Preparation Steps
For details, see the text. See also Table S1.

For each replicate of the Smart-seq protocol, we performed one run on the C1 platform from Fluidigm (Smart-seq/C1) using microfluidic chips that automatically capture up to 96 cells (Wu et al., 2014). We imaged captured cells, added lysis buffer together with the ERCCs, and we used the commercially available Smart-seq kit (Clontech) to generate full-length double-stranded cDNA that we converted into 96 sequencing libraries by tagmentation (Nextera, Illumina).

For each replicate of the Smart-seq2 protocol, we sorted mESCs by fluorescence activated cell sorting (FACS) into 96-well PCR plates containing lysis buffer and the ERCCs. We generated cDNA as described (Picelli et al., 2013, 2014b), and we used an in-house-produced Tn5 transposase (Picelli et al., 2014a) to generate 96 libraries by tagmentation. While Smart-seq/C1 and Smart-seq2 are very similar protocols that generate full-length libraries, they differ in how cells are isolated, their reaction volume, and in that the Smart-seq2 chemistry has been systematically optimized (Picelli et al., 2013, 2014b). The main disadvantage of both Smart-seq protocols is that the generation of full-length cDNA libraries precludes an early barcoding step and the incorporation of UMIs.

For each replicate of the SCRBS-seq protocol (Soumillon et al., 2014), we also sorted mESCs by FACS into 96-well PCR plates

containing lysis buffer and the ERCCs. Similar to the Smart-seq protocols, cDNA was generated by oligo-dT priming, template switching, and PCR amplification of full-length cDNA. However, the oligo-dT primers contained well-specific (i.e., cell-specific) barcodes and UMIs. Hence, cDNA from one plate could be pooled and then converted into sequencing libraries, using a modified tagmentation approach that enriches for the 3' ends. SCRBS-seq is optimized for small volumes and few handling steps.

The fourth method evaluated was Drop-seq, a recently developed microdroplet-based approach (Macosko et al., 2015). Here a flow of beads suspended in lysis buffer and a flow of a single-cell suspension were brought together in a microfluidic chip that generated nanoliter-sized emulsion droplets. On each bead, oligo-dT primers carrying a UMI and a unique, bead-specific barcode were covalently bound. Cells were lysed within these droplets, their mRNA bound to the oligo-dT-carrying beads, and, after breaking the droplets, cDNA and library generation was performed for all cells in parallel in one single tube. The ratio of beads to cells (20:1) ensured that the vast majority of beads had either no cell or one cell in its droplet. Hence, similar to SCRBS-seq, each cDNA molecule was labeled with a bead-specific (i.e., cell-specific) barcode and a UMI. We confirmed that

the Drop-seq protocol worked well in our setup by mixing mouse and human T cells, as recommended by [Macosko et al. \(2015\)](#) ([Figure S1A](#)). The main advantage of the protocol is that a high number of scRNA-seq libraries can be generated at low cost. One disadvantage of Drop-seq is that the simultaneous inclusion of ERCC spike-ins is quite expensive, as their addition would generate cDNA from ERCCs also in beads that have zero cells and thus would double the sequencing costs. As a proxy for the missing ERCC data, we used a published dataset ([Macosko et al., 2015](#)), where ERCC spike-ins were sequenced using the Drop-seq method without single-cell transcriptomes.

As a fifth method we chose CEL-seq2 ([Hashimshony et al., 2016](#)), an improved version of the original CEL-seq ([Hashimshony et al., 2012](#)) protocol, as implemented for microfluidic chips on Fluidigm's C1 ([Hashimshony et al., 2016](#)). As for Smart-seq/C1, this allowed us to capture 96 cells in two independent replicates and to include ERCCs in the cell lysis step. Similar to Drop-seq and SCRB-seq, cDNA was tagged with barcodes and UMIs; but, in contrast to the four PCR-based methods described above, CEL-seq2 relies on linear amplification by *in vitro* transcription after the initial reverse transcription. The amplified, bar-coded RNAs were harvested from the chip, pooled, fragmented, and reverse transcribed to obtain sequencing libraries.

MARS-seq, the sixth method evaluated, is a high-throughput implementation of the original CEL-seq method ([Jaitin et al., 2014](#)). In this protocol, cells were sorted by FACS in 384-well plates containing lysis buffer and the ERCCs. As in CEL-seq and CEL-seq2, amplified RNA with barcodes and UMIs were generated by *in vitro* transcription, but libraries were prepared on a liquid-handling platform. An overview of the methods and their workflows is provided in [Figure 2](#) and in [Table S1](#).

Processing of scRNA-Seq Data

For each method, we generated at least 48 libraries per replicate and sequenced between 241 and 866 million reads ([Figure 1](#); [Figure S1B](#)). All data were processed identically, with cDNA reads clipped to 45 bp and mapped using Spliced Transcripts Alignment to a Reference (STAR) ([Dobin et al., 2013](#)) and UMIs quantified using the Drop-seq pipeline ([Macosko et al., 2015](#)). To adjust for differences in sequencing depths, we selected all libraries with at least one million reads, and we downsampled them to one million reads each. This resulted in 96, 79, 73, 93, 162, and 187 libraries for CEL-seq2/C1, Drop-seq, MARS-seq, SCRB-seq, Smart-seq/C1, and Smart-seq2, respectively.

To exclude doublets (libraries generated from two or more cells) in the Smart-seq/C1 data, we analyzed microscope images and identified 16 reaction chambers with multiple cells. For the four UMI methods, we calculated the number of UMIs per library, and we found that libraries that have more than twice the mean total UMI count can be readily identified ([Figure S1C](#)). It is unclear whether these libraries were generated from two separate cells (doublets) or, for example, from one large cell before mitosis. However, for the purpose of this method comparison, we removed these three to nine libraries. To filter out low-quality libraries, we used a method that exploits the fact that transcript detection and abundance in low-quality libraries correlate poorly with high-quality libraries as well as with other low-quality libraries ([Petropoulos et al., 2016](#)). Therefore, we determined

the maximum Spearman correlation coefficient for each cell in all-to-all comparisons that allowed us to identify low-quality libraries as outliers of the distributions of correlation coefficients by visual inspection ([Figure S1D](#)). This filtering led to the removal of 21, 0, 4, 0, 16, and 30 cells for CEL-seq2/C1, Drop-seq, MARS-seq, SCRB-seq, Smart-seq/C1, and Smart-seq2, respectively.

In summary, we processed and filtered our data so that we ended up with a total of 583 high-quality scRNA-seq libraries that could be used for a fair comparison of the sensitivity, accuracy, precision, power, and efficiency of the methods.

Single-Cell Libraries Are Sequenced to a Reasonable Level of Saturation at One Million Reads

For all six methods, >50% of the reads could be unambiguously mapped to the mouse genome ([Figure 3A](#)), which is comparable to previous results ([Jaitin et al., 2014](#); [Wu et al., 2014](#)). Overall, between 48% (Smart-seq2) and 30% (Smart-seq/C1) of all reads were exonic and, thus, were used to quantify gene expression levels. However, the UMI data showed that only 14%, 5%, 7%, and 15% of the exonic reads were derived from independent mRNA molecules for CEL-seq2/C1, Drop-seq, MARS-seq, and SCRB-seq, respectively ([Figure 3A](#)). To quantify the relationship between the number of detected genes or mRNA molecules and the number of reads in more detail, we downsampled reads to varying depths, and we estimated to what extent libraries were sequenced to saturation ([Figure S2](#)). The number of unique mRNA molecules plateaued at 56,760 UMIs per library for CEL-seq2/C1 and 26,210 UMIs per library for MARS-seq, was still marginally increasing at 17,210 UMIs per library for Drop-seq, and was considerably increasing at 49,980 UMIs per library for SCRB-seq ([Figure S2C](#)). Notably, CEL-seq2/C1 and MARS-seq showed a steeper slope at low sequencing depths than both Drop-seq and SCRB-seq, potentially due to a less biased amplification by *in vitro* transcription. Hence, among the UMI methods, CEL-seq2/C1 and SCRB-seq libraries had the highest complexity of mRNA molecules, and this complexity was sequenced to a reasonable level of saturation with one million reads.

To investigate saturation also for non-UMI-based methods, we applied a similar approach at the gene level by counting the number of genes detected by at least one read. By fitting an asymptote to the downsampled data, we estimated that ~90% (Drop-seq and SCRB-seq) to 100% (CEL-seq2/C1, MARS-seq, Smart-seq/C1, and Smart-seq2) of all genes present in a library were detected at one million reads ([Figure 3B](#); [Figure S2A](#)). In particular, the deep sequencing of Smart-seq2 libraries showed clearly that the number of detected genes did not change when increasing the sequencing depth from one million to five million reads per cell ([Figure S2B](#)).

All in all, these analyses show that scRNA-seq libraries were sequenced to a reasonable level of saturation at one million reads, a cutoff that also has been suggested previously for scRNA-seq datasets ([Wu et al., 2014](#)). While it can be more efficient to invest in more cells at lower coverage (see our power analyses below), one million reads per cell is a reasonable sequencing depth for our purpose of comparing scRNA-seq methods.

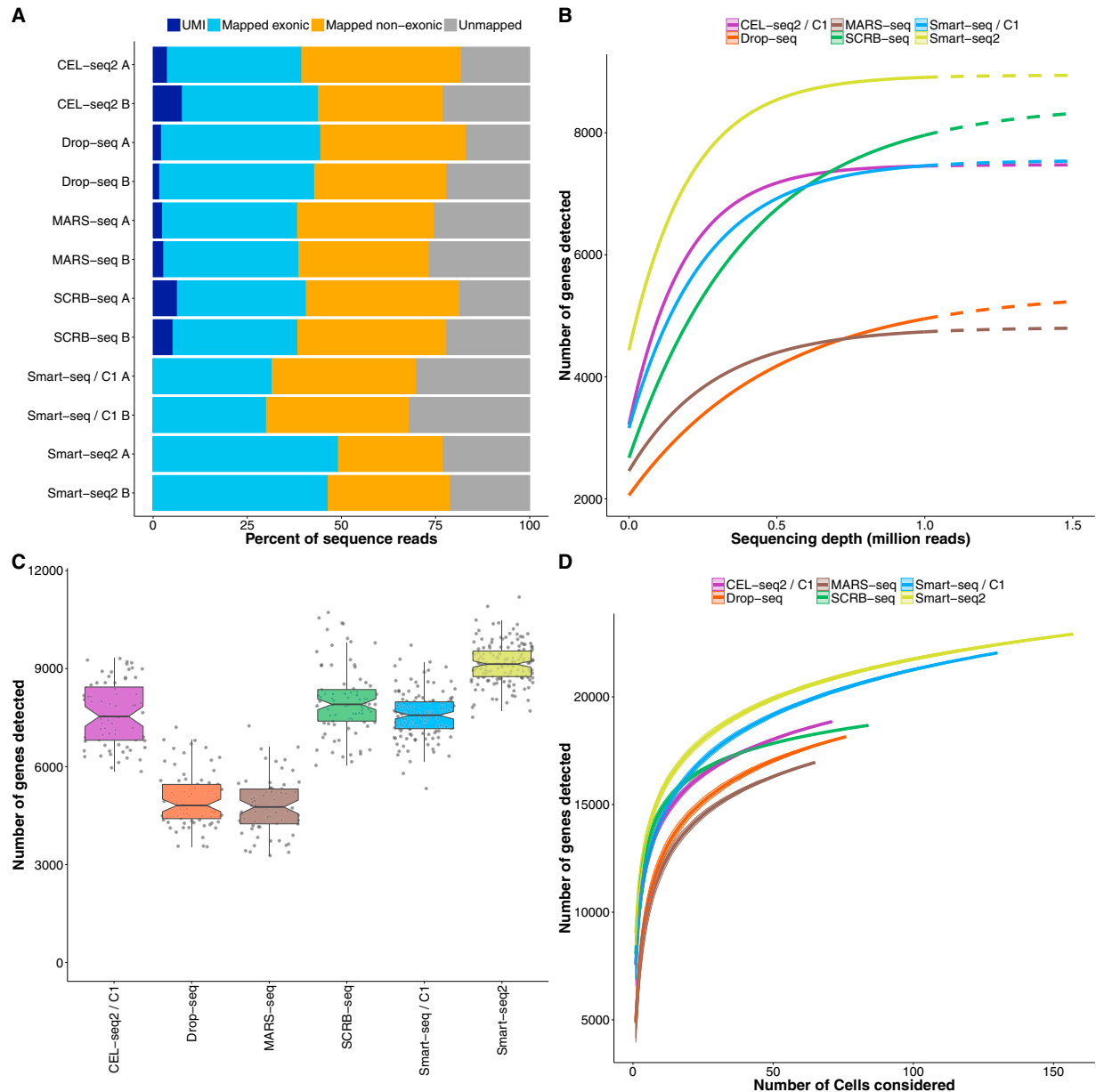


Figure 3. Sensitivity of scRNA-Seq Methods

(A) Percentage of reads (downsampled to one million per cell) that cannot be mapped to the mouse genome (gray) are mapped to regions outside exons (orange) or inside exons (blue). For UMI methods, dark blue denotes the exonic reads with unique UMIs.

(B) Median number of genes detected per cell (counts ≥ 1) when downsampling total read counts to the indicated depths. Dashed lines above one million reads represent extrapolated asymptotic fits.

(C) Number of genes detected (counts ≥ 1) per cell. Each dot represents a cell and each box represents the median and first and third quartiles per replicate and method.

(D) Cumulative number of genes detected as more cells are added. The order of cells considered was drawn randomly 100 times to display mean \pm SD (shaded area). See also Figures S3 and S4.

Smart-Seq2 Has the Highest Sensitivity

Taking the number of detected genes per cell as a measure of sensitivity, we found that Drop-seq and MARS-seq had the lowest

sensitivity, with a median of 4,811 and 4,763 genes detected per cell, respectively, while CEL-seq2/C1, SCRB-seq, and Smart-seq/C1 detected a median of 7,536, 7,906, and 7,572 genes per

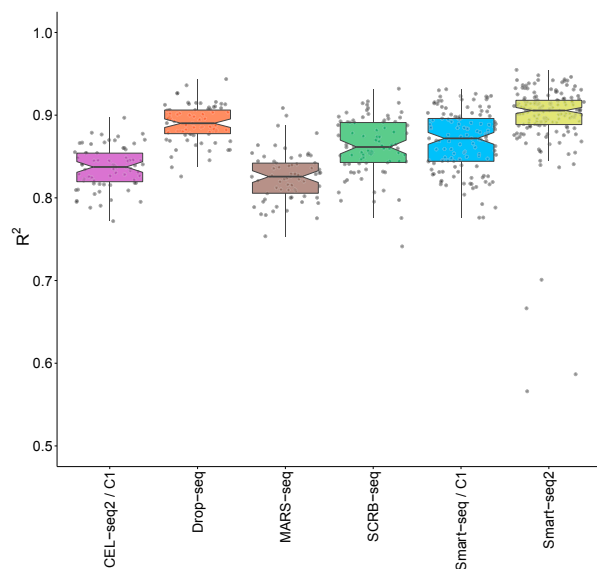


Figure 4. Accuracy of scRNA-Seq Methods

ERCC expression values (counts per million reads for Smart-seq/C1 and Smart-seq2 and UMIs per million reads for all others) were correlated to their annotated molarity. Shown are the distributions of correlation coefficients (adjusted R^2 of linear regression model) across methods. Each dot represents a cell/bead and each box represents the median and first and third quartiles. See also Figure S5.

cell (Figure 3C). Smart-seq2 detected the highest number of genes per cell with a median of 9,138. To compare the total number of genes detected across many cells, we pooled the sequence data of 65 cells per method, and we detected ~19,000 genes for CEL-seq2/C1, ~17,000 for MARS-seq, ~18,000 for Drop-seq and SCR-seq, ~20,000 for Smart-seq/C1, and ~21,000 for Smart-seq2 (Figure 3D). While the majority of genes (~13,000) were detected by all methods, ~400 genes were specific to each of the 3' counting methods, and ~1,000 genes were specific to each of the two full-length methods (Figure S3A). This higher sensitivity of both full-length methods also was apparent when plotting the genes detected in all available cells, as the 3' counting methods leveled off below 20,000 genes while the two full-length methods leveled off above 20,000 genes (Figure 3D). Such a difference could be caused by genes that have 3' ends that are difficult to map. However, we found that genes specific to Smart-seq2 and Smart-seq/C1 map as well to 3' ends as genes with similar length distribution that are not specifically detected by full-length methods (Figure S3B). Hence, it seems that full-length methods turn a slightly higher fraction of transcripts into sequenceable molecules than 3' counting methods and are more sensitive in this respect. Importantly, method-specific genes are detected in very few cells (87% of genes occur in one or two cells) with very low counts (mean counts < 0.2, Figure S3C). This suggests that they are unlikely to remain method specific at higher expression levels and that their impact on conclusions drawn from scRNA-seq data is rather limited (Lun et al., 2016).

Next, we investigated how reads are distributed along the mRNA transcripts for all genes. As expected, the 3' counting

methods showed a strong bias of reads mapped to the 3' end (Figure S3D). However, it is worth mentioning that a considerable fraction of reads also covered other segments of the transcripts, probably due to internal oligo-dT priming (Nam et al., 2002). Smart-seq2 showed a more even coverage than Smart-seq, confirming previous findings (Picelli et al., 2013). A general difference in expression values between 3' counting and full-length methods also was reflected in their strong separation by the first principal component, explaining 37% of the total variance, and when taking into account that one needs to normalize for gene length for the full-length methods (Figure S4E).

As an absolute measure of sensitivity, we compared the probability of detecting the 92 spiked-in ERCCs, for which the number of molecules available for library construction is known (Figures S4A and S4B). We determined the detection probability of each ERCC RNA as the proportion of cells with at least one read or UMI count for the particular ERCC molecule (Marinov et al., 2014). For Drop-seq, we used the previously published ERCC-only dataset (Macosko et al., 2015), and for the other five methods, 2%–5% of the one million reads per cell mapped to ERCCs that were sequenced to complete saturation at that level (Figure S5B). A 50% detection probability was reached at ~7, 11, 14, 16, 17, and 28 ERCC molecules for Smart-seq2, Smart-seq/C1, CEL-seq2/C1, SCR-seq, Drop-seq, and MARS-seq, respectively (Figure S4C). Notably, the sensitivity estimated from the number of detected genes does not fully agree with the comparison based on ERCCs. While Smart-seq2 was the most sensitive method in both cases, Drop-seq performed better and SCR-seq and MARS-seq performed worse when using ERCCs. The separate generation and sequencing of the Drop-seq ERCC libraries could be a possible explanation for their higher sensitivity. However, it remains unclear why SCR-seq and MARS-seq had a substantially lower sensitivity when using ERCCs. It has been noted before that ERCCs can be problematic for modeling endogenous mRNAs (Risso et al., 2014), potentially due to their shorter length, shorter poly-A tail, and their missing 5' cap (Grün and van Oudenaarden, 2015; Stegle et al., 2015). While ERCCs are still useful to gauge the absolute range of sensitivities, the thousands of endogenous mRNAs are likely to be a more reliable estimate for comparing sensitivities as we used the same cell type for all methods.

In summary, we find that Smart-seq2 is the most sensitive method, as it detects the highest number of genes per cell and the most genes in total across cells and has the most even coverage across transcripts. Smart-seq/C1 is slightly less sensitive per cell and detects almost the same number of genes across cells with slightly less even coverage. Among the 3' counting methods, CEL-seq2/C1 and SCR-seq detect about as many genes per cell as Smart-seq/C1, whereas Drop-seq and MARS-seq detect considerably fewer genes.

Accuracy of scRNA-Seq Methods

To measure the accuracy of transcript level quantifications, we compared the observed expression values (counts per million or UMIs per million) with the known concentrations of the 92 ERCC transcripts (Figure S5A). For each cell, we calculated the coefficient of determination (R^2) for a linear model fit (Figure 4). Methods differed significantly in their accuracy (Kruskal-Wallis

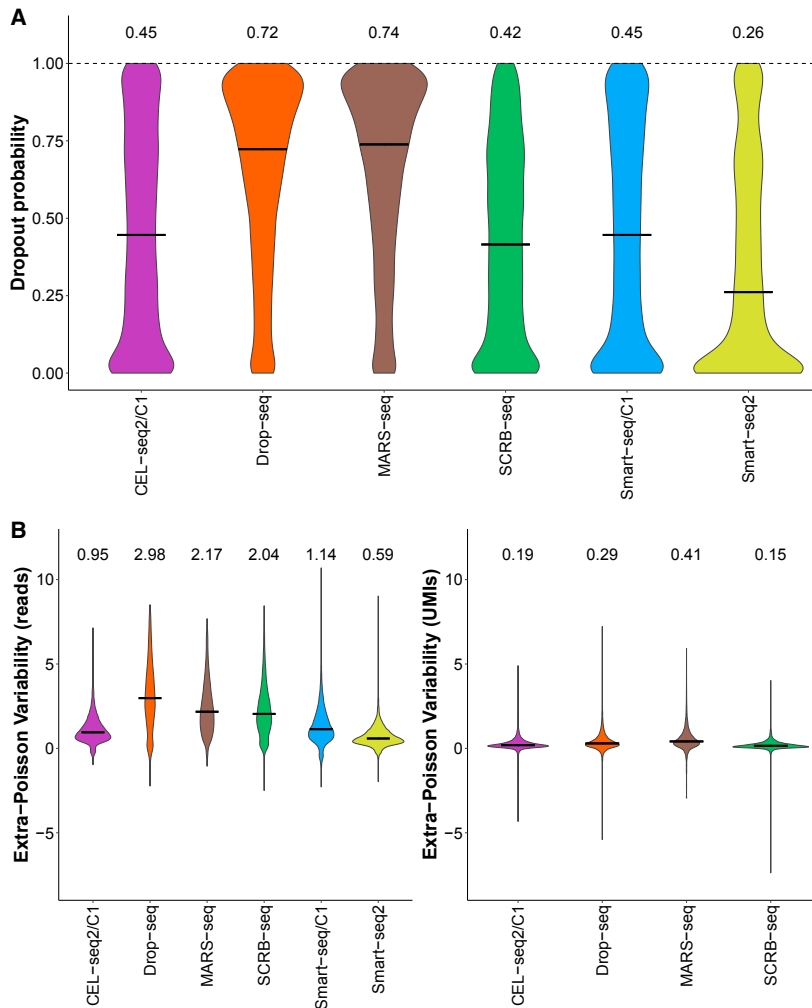


Figure 5. Precision of scRNA-Seq Methods

We compared precision among methods using the 13,361 genes detected in at least 25% of all cells by any method in a subsample of 65 cells per method.

(A) Distributions of dropout rates across the 13,361 genes are shown as violin plots, and medians are shown as bars and numbers.

(B) Extra Poisson variability across the 13,361 genes was calculated by subtracting the expected amount of variation due to Poisson sampling (square root of mean divided by mean) from the CV (SD divided by mean). Distributions are shown as violin plots and medians are shown as bars and numbers. For 349, 336, 474, 165, 201, and 146 genes for CEL-seq2/C1, Drop-seq, MARS-seq, SCRIB-seq, Smart-seq/C1, and Smart-seq2, respectively, no extra Poisson variability could be calculated. See also [Figures S6](#) and [S7](#).

the reproducibility of the expression-level estimate) is a major factor when choosing a method. As we used the same cell type under the same culture conditions for all methods, the amount of biological variation should be the same in the cells analyzed by each of the six methods. Hence, we can assume that differences in the total variation among methods are due to differences in their technical variation. Technical variation is substantial in scRNA-seq data primarily because a substantial fraction of mRNAs is lost during cDNA generation and small amounts of cDNA get amplified. Therefore, both the dropout probability and the amplification noise need to be considered when quantifying variation.

test, $p < 2.2e-16$), but all methods had a fairly high R^2 ranging between 0.83 (MARS-seq) and 0.91 (Smart-seq2). This suggests that, for all methods, transcript concentrations across this broad range can be predicted fairly well from expression values. As expected, accuracy was worse for narrower and especially for lower concentration ranges ([Figure S5C](#)). It is worth emphasizing that the accuracy assessed here refers to absolute expression levels across genes within cells. This accuracy can be important, for example, to identify marker genes with a high absolute mRNA expression level. However, the small differences in accuracy seen here will rarely be a decisive factor when choosing among the six protocols.

Precision of Amplified Genes Is Strongly Increased by UMIs

While a high accuracy is necessary to compare absolute expression levels, one of the most common experimental aims is to compare relative expression levels to identify differentially expressed genes or different cell types. Hence, the precision (i.e.,

Indeed, a mixture model including a dropout probability and a negative binomial distribution, modeling the overdispersion in the count data, have been shown to represent scRNA-seq data better than the negative binomial alone ([Finak et al., 2015](#); [Kharchenko et al., 2014](#)).

To compare precision without penalizing more sensitive methods, we selected a common set of 13,361 genes that were detected in 25% of the cells by at least one method ([Figure S6A](#)). We then analyzed these genes in a subsample of 65 cells per method to avoid a bias due to unequal numbers of cells. We estimated the dropout probability as the fraction of cells with zero counts ([Figure 5A](#); [Figure S6B](#)). As expected from the number of detected genes per cell ([Figure 3C](#)), MARS-seq had the highest median dropout probability (74%) and Smart-seq2 had the lowest (26%) ([Figure 5A](#)). To estimate the amplification noise of detected genes, we calculated the coefficient of variation (CV, SD divided by the mean, including zeros), and we subtracted the expected amount of variation due to Poisson sampling (i.e., the square root of the mean divided by the mean). This was possible

for 96.5% (MARS-seq) to 98.9% (Smart-seq2) of all the 13,361 genes. This extra Poisson variability includes biological variation (assumed to be the same across methods in our data) and technical variation, and the latter includes noise introduced by amplification (Brennecke et al., 2013; Grün et al., 2014; Stegle et al., 2015). That amplification noise can be a major factor is seen by the strong increase of extra Poisson variability when ignoring UMIs and considering read counts only (Figure 5B, left; Figure S7A). This is expected, as UMIs should remove amplification noise, which has been described previously for CEL-seq (Grün et al., 2014). For SCRB-seq and Drop-seq, which are PCR-based methods, UMIs removed even more extra Poisson variability than for CEL-seq2/C1 and MARS-seq (Figure 5B), which is in line with the notion that amplification by PCR is more noisy than amplification by *in vitro* transcription. Of note, Smart-seq2 had the lowest amplification noise when just considering reads (Figure 5B, left), potentially because its higher sensitivity requires less amplification and, hence, leads to less noise.

In summary, Smart-seq2 detects the common set of 13,361 genes in more cells than the UMI methods, but it has, as expected, more amplification noise than the UMI-based methods. How the different combinations of dropout rate and amplification noise affect the power of the methods is not evident, neither from this analysis nor from the total coefficient of variation that ignores the strong mean variance and mean dropout dependencies of scRNA-seq data (Figure S7B).

Power Is Determined by a Combination of Dropout Rates and Amplification Noise and Is Highest for SCRB-Seq

To estimate the combined impact of sensitivity and precision on the power to detect differential gene expression, we simulated scRNA-seq data given the observed dropout rates and variance for the 13,361 genes. As these depend strongly on the expression level of a gene, it is important to retain the mean variance and mean dropout relationships. To this end, we estimated the mean, the variance (i.e., the dispersion parameter of the negative binomial distribution), and the dropout rate for each gene and method. We then fitted a cubic smoothing spline to the resulting pairs of mean and dispersion estimates to predict the dispersion of a gene given its mean (Figure S8A). Furthermore, we applied a local polynomial regression model to account for the dropout probability given a gene's mean expression (Figure S8B). When simulating data according to these fits, we recovered distributions of dropout rates and variance closely matching the observed data (Figures S8C and S8D). To compare the power for differential gene expression among the methods, we simulated read counts for two groups of n cells and added log-fold changes to 5% of the 13,361 genes in one group. To mimic a biologically realistic scenario, these log-fold changes were drawn from observed differences between microglial subpopulations from a previously published dataset (Zeisel et al., 2015). Simulated datasets were tested for differential expression using limma (Ritchie et al., 2015), and the true positive rate (TPR) and the false discovery rate (FDR) were calculated. Of note, this does include undetected genes, i.e., the 2.5% (SCRB-seq) to 6.8% (MARS-seq) of the 13,361 genes that had fewer than two measurements in a particular method (Figure S6B) and for which we could not estimate the variance. In our simulations, these

genes could be drawn as differentially expressed, and in our TPR they were then counted as false negatives for the particular method. Hence, our power simulation framework considers the full range of dropout rates and is not biased against more sensitive methods.

First, we analyzed how the number of cells affects TPR and FDR by running 100 simulations each for a range of 16 to 512 cells per group (Figure 6A). FDRs were similar in all methods ranging from 3.9% to 8.7% (Figure S9A). TPRs differed considerably among methods and SCRB-seq performed best, reaching a median TPR of 80% with 64 cells. CEL-seq2/C1, Drop-seq, MARS-seq, and Smart-seq2 performed slightly worse, reaching 80% power with 86, 99, 110, and 95 cells per group, respectively, while Smart-seq/C1 needed 150 cells to reach 80% power (Figure 6A). When disregarding UMIs, Smart-seq2 performed best (Figure 6B), as expected from its low dropout rate and its low amplification noise when considering reads only (Figure 5B). Furthermore, power dropped especially for Drop-seq and SCRB-seq (Figure 6B), as expected from the strong increase in amplification noise of these two methods when considering reads only (Figure 5B). When we stratified our analysis (considering UMIs) across five bins of expression levels, the ranking of methods was recapitulated and showed that the lowest expression bin strongly limited the TPR in all methods (Figure S9B). This ranking also was recapitulated when we analyzed a set of 19 genes previously reported to contain cell-cycle variation in the 2i/LIF culture condition (Kolodziejczyk et al., 2015b). The variance of these cell-cycle genes was clearly higher than the variance of 19 pluripotency and housekeeping (ribosomal) genes in all methods. The p value of that difference was lowest for SCRB-seq, the most powerful method, and highest for Smart-seq/C1, the least powerful method (Figure S10D).

Notably, this power analysis, as well as the sensitivity, accuracy, and precision parameters analyzed above, includes the variation that is generated in the two technical replicates (batches) per method that we performed (Figure 1). These estimates were very similar among our technical replicates, and, hence, our method comparison is valid with respect to batch variations (Figures S10B–S10D). In addition, as batch effects are known to be highly relevant for interpreting scRNA-seq data (Hicks et al., 2015), we gauged the magnitude of batch effects with respect to identifying differentially expressed genes. To this end, we used limma to identify differentially expressed genes between batches (FDR < 1%), using 25 randomly selected cells per batch and method. All methods had significantly more genes differentially expressed between batches than expected from permutations (zero to four genes), with a median of 119 (Drop-seq) to ~1,135 (CEL-seq2/C1) differentially expressed genes (Figure S10A). Notably, genes were affected at random across methods, as there was no significant overlap among them (extended hypergeometric test [Kalinka, 2013], $p > 0.84$). Hence, this analysis once more emphasizes that batches are important to consider in the design of scRNA-seq experiments (Hicks et al., 2015). While a quantitative comparison of the magnitude of batch effects among methods would require substantially more technical replicates per method, the methods differ in their flexibility to incorporate batch effect into the experimental design, which is an important aspect to consider as discussed below.

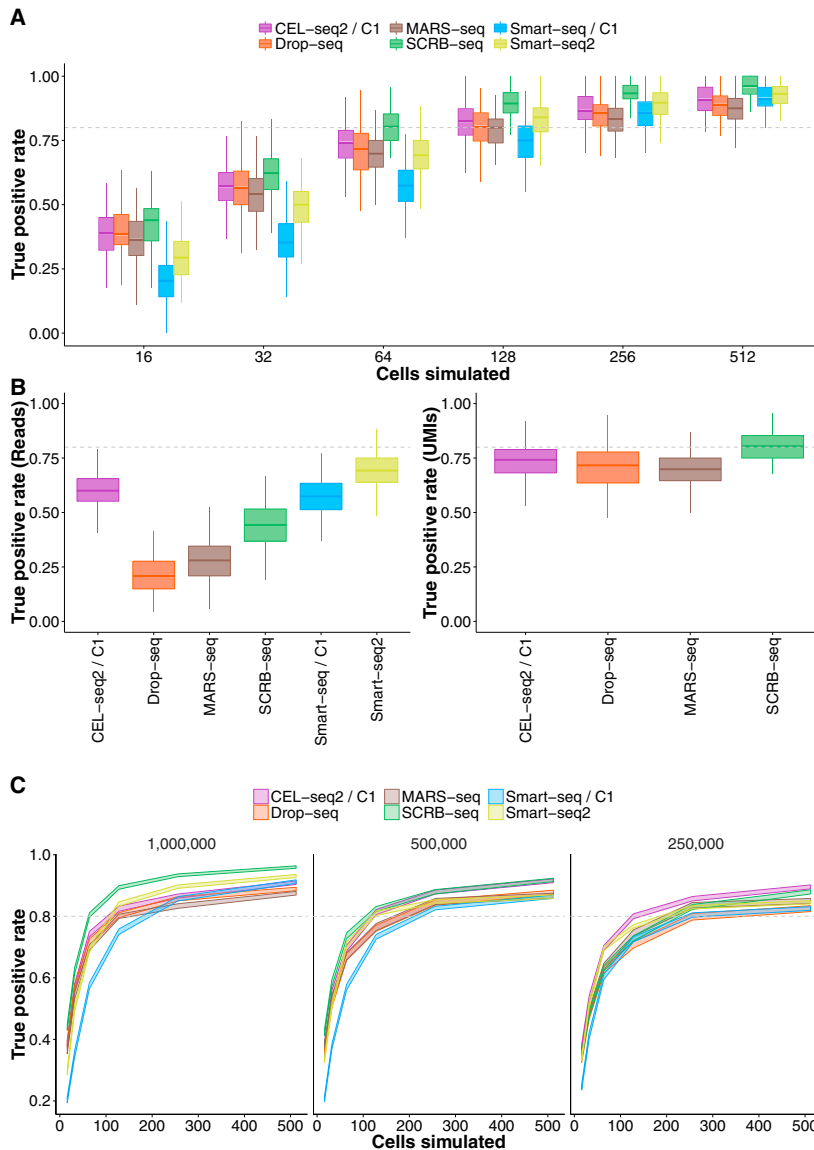


Figure 6. Power of scRNA-Seq Methods

Using the empirical mean/dispersion and mean/dropout relationships (Figures S8A and S8B), we simulated data for two groups of n cells each for which 5% of the 13,361 genes were differentially expressed, with log-fold changes drawn from observed differences between microglial subpopulations from a previously published dataset (Zeisel et al., 2015). The simulated data were then tested for differential expression using limma (Ritchie et al., 2015), from which the average true positive rate (TPR) and the average false discovery rate (FDR) were calculated (Figure S9A).

(A) TPR for one million reads per cell for sample sizes $n = 16$, $n = 32$, $n = 64$, $n = 128$, $n = 256$, and $n = 512$ per group. Boxplots represent the median and first and third quartiles of 100 simulations.

(B) TPR for one million reads per cell for $n = 64$ per group with and without using UMI information. Boxplots represent the median and first and third quartiles of 100 simulations.

(C) TPRs as in (A) using mean/dispersion and mean/dropout estimates from one million (as in A), 0.5 million, and 0.25 million reads. Line areas indicate the median power with SE from 100 simulations. See also Figures S8–S10 and Table 1.

including the scientific questions addressed, the experimental design, or the sample availability. However, the monetary cost is certainly an important one, and we used the results of our simulations to compare the costs among the methods for a given level of power.

Cost Efficiency Is Similarly High for Drop-Seq, MARS-Seq, SCR-seq, and Smart-Seq2

Given the number of cells needed to reach 80% power as simulated above for three sequencing depths (Figure 6C), we calculated the minimal costs to generate and sequence these libraries.

For example, at a sequencing depth of one million reads, SCR-seq requires 64 cells per group to reach 80% power. Generating 128 SCR-seq libraries costs ~260\$ and generating 128 million reads costs ~640\$. Note that the necessary paired-end reads for CEL-seq2/C1, SCR-seq, MARS-seq, and Drop-seq can be generated using a 50-cycle sequencing kit, and, hence, we assume that sequencing costs are the same for all methods.

Calculating minimal costs this way, Drop-seq (690\$) is the most cost-effective method when sequencing 254 cells at a depth of 250,000 reads, and SCR-seq (810\$), MARS-seq (820\$), and Smart-seq2 (1,090\$) are slightly more expensive at the same performance (Table 1). For Smart-seq2 it should be stressed that the use of in-house-produced Tn5 transposase (Picelli et al., 2014a) is required to keep the cost at this level, as

As a next step, we analyzed how the performance of the six methods depends on sequencing depth. To this end, we performed power simulations as above, but we estimated the mean dispersion and mean dropout relationships from data downsampled to 500,000 or 250,000 reads per cell. Overall, the decrease in power was moderate (Figure 6C; Table 1) and followed the drop in sensitivity at different sequencing depths (Figure 3B). While Smart-seq2 and CEL-seq2/C1 needed just 1.3-fold more cells at 0.25 million reads than at one million reads to reach 80% power, SCR-seq and Drop-seq required 2.6-fold more cells (Table 1). In summary, SCR-seq is the most powerful method at one million reads and half a million reads, but CEL-seq2/C1 is the most powerful method at a sequencing depth of 250,000 reads. The optimal balance between the number of cells and their sequencing depth depends on many factors,

Table 1. Cost Efficiency Extrapolation for Single-Cell RNA-Seq Experiments

Method	TPR ^a	FDR ^a (%)	Cell per Group ^b	Library Cost (\$)	Minimal Cost ^c (\$)
CEL-seq2/C1	0.8	~6.1	86/100/110	~9	~2,420/2,310/2,250
Drop-seq	0.8	~8.4	99/135/254	~0.1	~1,010/700/690
MARS-seq	0.8	~7.3	110/135/160	~1.3	~1,380/1,030/820
SCRB-seq	0.8	~6.1	64/90/166	~2	~900/810/1,080
Smart-seq/C1	0.8	~4.9	150/172/215	~25	~9,010/9,440/11,290
Smart-seq2 (commercial)	0.8	~5.2	95/105/128	~30	~10,470/11,040/13,160
Smart-seq2 (in-house Tn5)	0.8	~5.2	95/105/128	~3	~1,520/1,160/1,090

See also Figure 6.

^aTrue positive rate and false discovery rate are based on simulations (Figure 6; Figure S9).

^bSequencing depth of one, 0.5, and 0.25 million reads.

^cAssuming \$5 per one million reads.

was done in our experiments. When instead using the Tn5 transposase of the commercial Nextera kit as described (Picelli et al., 2014b), the costs for Smart-seq2 are 10-fold higher. Even if one reduces the amount of Nextera transposase to a quarter, as done in the Smart-seq/C1 protocol, the Smart-seq2 protocol is still four times more expensive than the early barcoding methods. CEL-seq2/C1 is fairly expensive due to the microfluidic chips that make up 69% of the library costs, and Smart-seq/C1 is almost 13-fold less efficient than Drop-seq due to its high library costs that arise from the microfluidic chips, the commercial Smart-seq kit, and the costs for commercial Nextera XT kits.

Of note, these calculations are the minimal costs of the experiment and several factors are not considered, such as labor costs, costs to set up the methods, costs to isolate cells of interest, or costs due to practical constraints in generating a fixed number of scRNA-seq libraries with a fixed number of reads. In many experimental settings, independent biological and/or technical replicates are needed when investigating particular factors, such as genotypes or developmental time points, and Smart-seq/C1, CEL-seq2/C1, and Drop-seq are less flexible in distributing scRNA-seq libraries across replicates than the other three methods that use PCR plates. Furthermore, the costs are increased by unequal sampling from the included cells as well as from sequencing reads from cells that are excluded. In our case, between 6% (SCRB-seq) and 32% (Drop-seq) of the reads came from cell barcodes that were not included. While it is difficult to exactly calculate and compare these costs among methods, it is clear that they will increase the costs for Drop-seq relatively more than for the other methods. In summary, we find that Drop-seq, SCRIB-seq, and MARS-seq are the most cost-effective methods, closely followed by Smart-seq2, if using an in-house-produced transposase.

DISCUSSION

Here we have provided an in-depth comparison of six prominent scRNA-seq protocols. To this end, we generated data for all six compared methods from the same cells, cultured under the same condition in the same laboratory. While there would be many more datasets and methods for a comparison of the sensitivity and accuracy of the ERCCs (Svensson et al., 2016), our approach provides a more controlled and comprehensive com-

parison across thousands of endogenous genes. This is important, as can be seen by the different sensitivity estimates that we obtained for Drop-seq, MARS-seq, and SCRIB-seq using the ERCCs. In our comparison, we clearly find that Smart-seq2 is the most sensitive method, closely followed by SCRIB-seq, Smart-seq/C1, and CEL-seq2/C1, while Drop-seq and MARS-seq detect nearly 50% fewer genes per cell (Figures 3B and 3C). In addition, Smart-seq2 shows the most even read coverage across transcripts (Figure S3D), making it the most appropriate method for the detection of alternative splice forms and for analyses of allele-specific expression using SNPs (Deng et al., 2014; Reinis et al., 2016). Hence, Smart-seq2 is certainly the most suitable method when an annotation of single-cell transcriptomes is the focus. Furthermore, we find that Smart-seq2 is also the most accurate method (i.e., it has the highest correlation of known ERCC spike-in concentrations and read counts per million), which is probably related to its higher sensitivity. Hence, differences in expression values across transcripts within the same cell predict differences in the actual concentrations of these transcripts well. All methods do this rather well, at least for higher expression levels, and we think that the small differences among methods will rarely be a decisive factor. Importantly, the accuracy of estimating transcript concentrations across cells (relevant, e.g., for comparing the total RNA content of cells) depends on different factors and cannot be compared well among the tested methods as it would require known concentration differences of transcripts across cells. However, it is likely that methods that can use UMIs and ERCCs (CEL-seq2/C1, MARS-seq, and SCRIB-seq) would have a strong advantage in this respect.

How well relative expression levels of the same genes can be compared across cells depends on two factors. First, how often (i.e., in how many cells and from how many molecules) it is measured. Second, with how much technical variation (i.e., with how much noise, e.g., from amplification) it is measured. For the first factor (dropout probability), we find Smart-seq2 to be the best method (Figure 5A), as expected from its high gene detection sensitivity. For the second factor (extra Poisson variability), we find the four UMI methods to perform better (Figure 5B), as expected from their ability to eliminate variation introduced by amplification. To assess the combined effect of these two factors, we performed simulations for differential gene

expression scenarios (Figure 6). This allowed us to translate the sensitivity and precision parameters into the practically relevant power to detect differentially expressed genes. Of note, our power estimates include the variation that is caused by the two different replicates per method that constitutes an important part of the variation. Our simulations show that, at a sequencing depth of one million reads, SCRB-seq has the highest power, probably due to a good balance of high sensitivity and low amplification noise. Furthermore, amplification noise and power strongly depend on the use of UMIs, especially for the PCR-based methods (Figures 5B and 6B; Figure S7). Notably, this is due to the large amount of amplification needed for scRNA-seq libraries, as the effect of UMIs on power for bulk RNA-seq libraries is negligible (Parekh et al., 2016).

Perhaps practically most important, our power simulations also allow us to compare the efficiency of the methods by calculating the costs to generate the data for a given level of power. Using minimal cost calculations, we find that Drop-seq is the most cost-effective method, closely followed by SCRB-seq, MARS-seq, and Smart-seq2. However, Drop-seq costs are likely to be more underestimated, due to lower flexibility in generating a specified number of libraries and the higher fraction of reads that come from bad cells. Hence, all four UMI methods are in practice probably similarly cost-effective. In contrast, for Smart-seq2 to be similarly cost-effective it is absolutely necessary to use in-house-produced transposase or to drastically reduce volumes of commercial transposase kits (Lamble et al., 2013; Mora-Castilla et al., 2016).

Given comparable efficiencies of Drop-seq, MARS-seq, SCRB-seq, and Smart-seq2, additional factors will play a role when choosing a suitable method for a particular question. Due to its low library costs, Drop-seq is probably preferable when analyzing large numbers of cells at low coverage (e.g., to find rare cell types). On the other hand, Drop-seq in its current setup requires a relatively large amount of cells (>6,500 for 1 min of flow). Hence, if few and/or unstable cells are isolated by FACS, the SCRB-seq, MARS-seq, or Smart-seq2 protocols are probably preferable. Additional advantages of these methods over Drop-seq include that technical variation can be estimated from ERCCs for each cell, which can be helpful to estimate biological variation (Kim et al., 2015; Vallejos et al., 2016), and that the exact same setup can be used to generate bulk RNA-seq libraries. While SCRB-seq is slightly more cost-effective than MARS-seq and has the advantage that one does not need to produce the transposase in-house, Smart-seq2 is preferable when transcriptome annotation, identification of sequence variants, or the quantification of different splice forms is of interest. Furthermore, the presence of batch effects shows that experiments need to be designed in a way that does not confound batches with biological factors (Hicks et al., 2015). Practically, plate-based methods might currently accommodate complex experimental designs with various biological factors more easily than microfluidic chips.

We find that Drop-seq, MARS-seq, SCRB-seq, and Smart-seq2 (using in-house transposase) are 2- to 13-fold more cost efficient than CEL-seq2/C1, Smart-seq/C1, and Smart-seq2 (using commercial transposase). Hence, the latter methods

would need to increase in their power and/or decrease in their costs to be competitive. The efficiency of the Fluidigm C1 platform can be further increased by microfluidic chips with a higher throughput, as available in the high-throughput (HT) mRNA-seq integrated fluidic circuit (IFC) chip. While CEL-seq2/C1 has been found to more sensitive than the plate-based version of CEL-seq2 (Hashimshony et al., 2016), the latter might be more efficient when considering its lower costs. Our finding that Smart-seq2 is the most sensitive protocol also hints toward further possible improvements of SCRB-seq and Drop-seq. As these methods also rely on template switching and PCR amplification, the improvements found in the systematic optimization of Smart-seq2 (Picelli et al., 2013) also could improve the sensitivity of SCRB-seq and Drop-seq. Furthermore, the costs of SCRB-seq libraries per cell can be halved when switching to a 384-well format (Soumillon et al., 2014). Similarly, improvements made for CEL-seq2 (Hashimshony et al., 2016) could be incorporated into the MARS-seq protocol. Hence, it is clear that scRNA-seq protocols will become even more efficient in the future. The results of our comparative analyses of six currently prominent scRNA-seq methods may facilitate such developments, and they provide a framework for method evaluation in the future.

In summary, we systematically compared six prominent scRNA-seq methods and found that Drop-seq is preferable when quantifying transcriptomes of large numbers of cells with low sequencing depth, SCRB-seq and MARS-seq is preferable when quantifying transcriptomes of fewer cells, and Smart-seq2 is preferable when annotating and/or quantifying transcriptomes of fewer cells as long one can use in-house-produced transposase. Our analysis allows an informed choice among the tested methods, and it provides a framework for benchmarking future improvements in scRNA-seq methodologies.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
 - Published data
 - Single cell RNA-seq library preparations
 - DNA sequencing
- QUANTIFICATION AND STATISTICAL ANALYSIS
 - Basic data processing and sequence alignment
 - Power Simulations
 - ERCC capture efficiency
 - Cost efficiency calculation
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes ten figures and one table and can be found with this article online at <http://dx.doi.org/10.1016/j.molcel.2017.01.023>.

AUTHOR CONTRIBUTIONS

C.Z. and W.E. conceived the experiments. C.Z. prepared scRNA-seq libraries and analyzed the data. B.V. implemented the power simulation framework and estimated the ERCC capture efficiencies. S.P. helped in data processing and power simulations. B.R. prepared the Smart-seq2 scRNA-seq libraries. A.G.-A. and H.H. established and performed the MARS-seq library preps. M.S. performed the cell culture of mESCs. W.E. and H.L. supervised the experimental work and I.H. provided guidance in data analysis. C.Z., I.H., B.R., and W.E. wrote the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

We thank Rickard Sandberg for facilitating the Smart-seq2 sequencing. We thank Christopher Mulholland for assistance with FACS, Dominik Alterauge for help establishing the Drop-seq method, and Stefan Krebs and Helmut Blum from the LAFUGA platform for sequencing. We are grateful to Magali Soumilion and Tarjei Mikkelsen for providing the SCRB-seq protocol. This work was supported by the Deutsche Forschungsgemeinschaft (DFG) through LMUexcellent and the SFB1243 (Subproject A01/A14/A15) as well as a travel grant to C.Z. by the Boehringer Ingelheim Fonds.

Received: August 8, 2016

Revised: December 1, 2016

Accepted: January 17, 2017

Published: February 9, 2017

REFERENCES

- Brennecke, P., Anders, S., Kim, J.K., Kolodziejczyk, A.A., Zhang, X., Proserpio, V., Baying, B., Benes, V., Teichmann, S.A., Marioni, J.C., and Heisler, M.G. (2013). Accounting for technical noise in single-cell RNA-seq experiments. *Nat. Methods* **10**, 1093–1095.
- Deng, Q., Ramsköld, D., Reinius, B., and Sandberg, R. (2014). Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science* **343**, 193–196.
- Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21.
- ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74.
- Finak, G., McDavid, A., Yajima, M., Deng, J., Gersuk, V., Shalek, A.K., Slichter, C.K., Miller, H.W., McElrath, M.J., Pric, M., et al. (2015). MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* **16**, 278.
- Frazee, A.C., Jaffe, A.E., Langmead, B., and Leek, J.T. (2015). Polyester: simulating RNA-seq datasets with differential transcript expression. *Bioinformatics* **31**, 2778–2784.
- Gokce, O., Stanley, G.M., Treutlein, B., Neff, N.F., Camp, J.G., Malenka, R.C., Rothwell, P.E., Fuccillo, M.V., Südhof, T.C., and Quake, S.R. (2016). Cellular taxonomy of the mouse striatum as revealed by single-cell RNA-seq. *Cell Rep.* **16**, 1126–1137.
- Grün, D., and van Oudenaarden, A. (2015). Design and analysis of single-cell sequencing experiments. *Cell* **163**, 799–810.
- Grün, D., Kester, L., and van Oudenaarden, A. (2014). Validation of noise models for single-cell transcriptomics. *Nat. Methods* **11**, 637–640.
- Hashimshony, T., Wagner, F., Sher, N., and Yanai, I. (2012). CEL-seq: single-cell RNA-seq by multiplexed linear amplification. *Cell Rep.* **2**, 666–673.
- Hashimshony, T., Senderovich, N., Avital, G., Klochendler, A., de Leeuw, Y., Anavy, L., Gennert, D., Li, S., Livak, K.J., Rozenblatt-Rosen, O., et al. (2016). CEL-seq2: sensitive highly-multiplexed single-cell RNA-Seq. *Genome Biol.* **17**, 77.
- Hicks, S.C., Teng, M., and Irizarry, R.A. (2015). On the widespread and critical impact of systematic bias and batch effects in single-cell RNA-Seq data. *bioRxiv*. <http://dx.doi.org/10.1101/025528>.
- Islam, S., Zeisel, A., Joost, S., La Manno, G., Zajac, P., Kasper, M., Lönnerberg, P., and Linnarsson, S. (2014). Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat. Methods* **11**, 163–166.
- Jaitin, D.A., Kenigsberg, E., Keren-Shaul, H., Elefant, N., Paul, F., Zaretzky, I., Mildner, A., Cohen, N., Jung, S., Tanay, A., and Amit, I. (2014). Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* **343**, 776–779.
- Jiang, L., Schlesinger, F., Davis, C.A., Zhang, Y., Li, R., Salit, M., Gingeras, T.R., and Oliver, B. (2011). Synthetic spike-in standards for RNA-seq experiments. *Genome Res.* **21**, 1543–1551.
- Kalinka, A.T. (2013). The probability of drawing intersections: extending the hypergeometric distribution. *arXiv*, arXiv:1305.0717. <https://arxiv.org/abs/1305.0717>.
- Kharchenko, P.V., Silberstein, L., and Scadden, D.T. (2014). Bayesian approach to single-cell differential expression analysis. *Nat. Methods* **11**, 740–742.
- Kim, J.K., Kolodziejczyk, A.A., Illicic, T., Teichmann, S.A., and Marioni, J.C. (2015). Characterizing noise structure in single-cell RNA-seq distinguishes genuine from technical stochastic allelic expression. *Nat. Commun.* **6**, 8687.
- Kivioja, T., Vähärautio, A., Karlsson, K., Bonke, M., Enge, M., Linnarsson, S., and Taipale, J. (2011). Counting absolute numbers of molecules using unique molecular identifiers. *Nat. Methods* **9**, 72–74.
- Klein, A.M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D.A., and Kirschner, M.W. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187–1201.
- Kolodziejczyk, A.A., Kim, J.K., Svensson, V., Marioni, J.C., and Teichmann, S.A. (2015a). The technology and biology of single-cell RNA sequencing. *Mol. Cell* **58**, 610–620.
- Kolodziejczyk, A.A., Kim, J.K., Tsang, J.C.H., Illicic, T., Henriksson, J., Natarajan, K.N., Tuck, A.C., Gao, X., Bühler, M., Liu, P., et al. (2015b). Single cell RNA-sequencing of pluripotent states unlocks modular transcriptional variation. *Cell Stem Cell* **17**, 471–485.
- Lamble, S., Batty, E., Attar, M., Buck, D., Bowden, R., Lunter, G., Crook, D., El-Fahmawi, B., and Piazza, P. (2013). Improved workflows for high throughput library preparation using the transposome-based Nextera system. *BMC Biotechnol.* **13**, 104.
- Law, C.W., Chen, Y., Shi, W., and Smyth, G.K. (2014). voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29.
- Li, E., Bestor, T.H., and Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915–926.
- Liao, Y., Smyth, G.K., and Shi, W. (2013). The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res.* **41**, e108.
- Lun, A.T.L., Bach, K., and Marioni, J.C. (2016). Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. *Genome Biol.* **17**, 75.
- Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214.
- Marinov, G.K., Williams, B.A., McCue, K., Schroth, G.P., Gertz, J., Myers, R.M., and Wold, B.J. (2014). From single-cell to cell-pool transcriptomes: stochasticity in gene expression and RNA splicing. *Genome Res.* **24**, 496–510.
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12.
- Mora-Castilla, S., To, C., Vaezeslami, S., Morey, R., Srinivasan, S., Dumdie, J.N., Cook-Andersen, H., Jenkins, J., and Laurent, L.C. (2016). Miniaturization technologies for efficient single-cell library preparation for next-generation sequencing. *J. Lab. Autom.* **21**, 557–567.

- Nam, D.K., Lee, S., Zhou, G., Cao, X., Wang, C., Clark, T., Chen, J., Rowley, J.D., and Wang, S.M. (2002). Oligo(dT) primer generates a high frequency of truncated cDNAs through internal poly(A) priming during reverse transcription. *Proc. Natl. Acad. Sci. USA* **99**, 6152–6156.
- Parekh, S., Ziegenhain, C., Vieth, B., Enard, W., and Hellmann, I. (2016). The impact of amplification on differential expression analyses by RNA-seq. *Sci. Rep.* **6**, 25533.
- Petropoulos, S., Edsgård, D., Reinius, B., Deng, Q., Panula, S.P., Codeluppi, S., Plaza Reyes, A., Linnarsson, S., Sandberg, R., and Lanner, F. (2016). Single-cell RNA-seq reveals lineage and X chromosome dynamics in human preimplantation embryos. *Cell* **165**, 1012–1026.
- Picelli, S., Björklund, Å.K., Faridani, O.R., Sagasser, S., Winberg, G., and Sandberg, R. (2013). Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* **10**, 1096–1098.
- Picelli, S., Björklund, Å.K., Reinius, B., Sagasser, S., Winberg, G., and Sandberg, R. (2014a). Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res.* **24**, 2033–2040.
- Picelli, S., Faridani, O.R., Björklund, Å.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014b). Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9**, 171–181.
- Reinius, B., Mold, J.E., Ramsköld, D., Deng, Q., Johnsson, P., Michaëlsson, J., Frisén, J., and Sandberg, R. (2016). Analysis of allelic expression patterns in clonal somatic cells by single-cell RNA-seq. *Nat. Genet.* **48**, 1430–1435.
- Renaud, G., Stenzel, U., Maricic, T., Wiebe, V., and Kelso, J. (2015). deML: robust demultiplexing of Illumina sequences using a likelihood-based approach. *Bioinformatics* **31**, 770–772.
- Risso, D., Ngai, J., Speed, T.P., and Dudoit, S. (2014). Normalization of RNA-seq data using factor analysis of control genes or samples. *Nat. Biotechnol.* **32**, 896–902.
- Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47.
- Soumillon, M., Cacchiarelli, D., Semrau, S., van Oudenaarden, A., and Mikkelsen, T.S. (2014). Characterization of directed differentiation by high-throughput single-cell RNA-seq. *bioRxiv*. <http://dx.doi.org/10.1101/003236>.
- Stegle, O., Teichmann, S.A., and Marioni, J.C. (2015). Computational and analytical challenges in single-cell transcriptomics. *Nat. Rev. Genet.* **16**, 133–145.
- Svensson, V., Natarajan, K.N., Ly, L.-H., Miragaia, R.J., Labalette, C., Macaulay, I.C., Cvejic, A., and Teichmann, S.A. (2016). Power analysis of single cell RNA-sequencing experiments. *bioRxiv*. <http://dx.doi.org/10.1101/073692>.
- Tirosh, I., Izar, B., Prakadan, S.M., Wadsworth, M.H., 2nd, Treacy, D., Trombetta, J.J., Rotem, A., Rodman, C., Lian, C., Murphy, G., et al. (2016). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196.
- Vallejos, C.A., Richardson, S., and Marioni, J.C. (2016). Beyond comparisons of means: understanding changes in gene expression at the single-cell level. *Genome Biol.* **17**, 70.
- Wagner, A., Regev, A., and Yosef, N. (2016). Revealing the vectors of cellular identity with single-cell genomics. *Nat. Biotechnol.* **34**, 1145–1160.
- Wu, A.R., Neff, N.F., Kalisky, T., Dalerba, P., Treutlein, B., Rothenberg, M.E., Mburu, F.M., Mantalas, G.L., Sim, S., Clarke, M.F., and Quake, S.R. (2014). Quantitative assessment of single-cell RNA-sequencing methods. *Nat. Methods* **11**, 41–46.
- Zeisel, A., Muñoz-Manchado, A.B., Codeluppi, S., Lönnerberg, P., La Manno, G., Jureus, A., Marques, S., Munguba, H., He, L., Betsholtz, C., et al. (2015). Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **347**, 1138–1142.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, Peptides, and Recombinant Proteins		
Esgro recombinant mouse LIF	Millipore	ESG1107
CHIR99021	Axon Med Chem	1386
PD0325901	Axon Med Chem	1408
2-Mercaptoethanol	Sigma-Aldrich	M3148
FBS	Sigma-Aldrich	F7524
Penicillin/Streptomycin	Sigma-Aldrich	P4333
MEM non-essential amino acids	Sigma-Aldrich	M7145
L-glutamine	Sigma-Aldrich	G7513
Dulbecco's modified Eagle's medium	Sigma-Aldrich	D6429
Perfluorooctanol	Sigma-Aldrich	370533
Maxima H- Reverse Transcriptase	Thermo Fisher Scientific	EP0753
SuperScript II	Life Technologies	18064071
Exonuclease I	New England Biolabs	M0293L
RNAprotect Cell Reagent	QIAGEN	76526
RNase inhibitor	Promega	N2515
RNase inhibitor	Lucigen	30281-2-LU
Phusion HF buffer	New England Biolabs	B0518S
Proteinase K	Ambion	AM2546
KAPA HiFi HotStart polymerase	KAPA Biosystems	KAPBKK2602
Phusion HF PCR Master Mix	Thermo Fisher Scientific	F531L
dNTPs	New England Biolabs	N0447L
Triton X-100	Sigma-Aldrich	T8787
SDS	Sigma-Aldrich	L3771
Tn5 transposase	Picelli et al., 2014a	N/A
Critical Commercial Assays		
C1 Single-Cell System	Fluidigm	N/A
C1 IFC for Open App (10-17 μ m)	Fluidigm	100-8134
C1 IFC for mRNA-seq (10-17 μ m)	Fluidigm	100-6041
Nextera XT DNA Sample Preparation Kit	Illumina	FC-131-1096
SMARTer Ultra Low RNA Kit for Fluidigm C1	Clontech	634833
MinElute Gel Extraction Kit	QIAGEN	28606
Deposited Data		
single-cell RNA-seq data	This paper	GEO: GSE75790
Drop-seq ERCC data	Macosko et al., 2015	GEO: GSE66694
Experimental Models: Cell Lines		
J1 mouse embryonic stem cells	Li et al., 1992	N/A
Sequence-Based Reagents		
Nextera XT Index Kit	Illumina	FC-121-1012
SCRB-seq P5 primer, AATGATACGGCGACCACCG AGATCTACACTCTTTCCCTACACGACGCTCTTC CG*A*T*C*T, * PTO bond	IDT	N/A
SCRB-seq oligo-dT primer, Biotin-ACACTCTTTCCCT ACACGACGCTCTTCGATCT[BC6][N10][T30]VN	IDT	"TruGrade Ultramer"

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
SCRB-seq template-switch oligo, iCiGiCACACTCTTTCC CTACACGACGCrGrGrG	Eurogentech	N/A
Drop-seq P5 primer, AATGATACGGCGACCACCGAGA TCTACACGCCT GTCCGCGGAAGCAGTGGTATCAACG CAGAGT*A*C, * PTO bond	IDT	N/A
Drop-seq oligo-dT primer beads, Bead-Linker- TTTTTTTAAGCAGTGGTATCAAC GCAGAGTAC[BC12][N8][T30]	Chemgenes	MACOSKO-2011-10
Drop-seq template-switch oligo, AAGCAGTGGTATCA ACGCAGAGTGAATrGrGrG	IDT	N/A
CEL-seq2 oligo-dT primer, GCCGGTAATACGACTCACTATA GGGAGTTCTACAGTCCGACGATC[N6][BC6][T25]	Sigma-Aldrich	N/A
ERCC RNA Spike-In Mix	Ambion	4456740
Software and Algorithms		
STAR	Dobin et al., 2013	https://github.com/alexdobin/STAR
Drop-seq tools	Macosko et al., 2015	http://mccarrolllab.com/dropseq/
featureCounts	Liao et al., 2013	https://bioconductor.org/packages/release/bioc/html/Rsubread.html
R	N/A	www.r-project.org
Other		
Drop-seq PDMS device	Nanoshift	Drop-seq
2% E-Gel Agarose EX Gels	Life Technologies	G402002

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the corresponding author Wolfgang Enard (enard@biologie.uni-muenchen.de).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

J1 mouse embryonic stem cells ([Li et al., 1992](#)) were maintained on gelatin-coated dishes in Dulbecco's modified Eagle's medium supplemented with 16% fetal bovine serum (FBS, Sigma-Aldrich), 0.1 mM β -mercaptoethanol (Sigma-Aldrich), 2 mM L-glutamine, 1x MEM non-essential amino acids, 100 U/ml penicillin, 100 μ g/ml streptomycin (Sigma-Aldrich), 1000 U/ml recombinant mouse LIF (Millipore) and 2i (1 μ M PD032591 and 3 μ M CHIR99021 (Axon Medchem, Netherlands). J1 embryonic stem cells were obtained from E. Li and T. Chen and mycoplasma free determined by a PCR-based test. Cell line authentication was not recently performed.

METHOD DETAILS**Published data**

Drop-seq ERCC ([Macosko et al., 2015](#)) data were obtained under accession GEO: GSE66694. Raw fastq files were extracted using the SRA toolkit (2.3.5). We trimmed cDNA reads to the same length and processed raw reads in the same way as data sequenced for this study.

Single cell RNA-seq library preparations**CEL-seq2/C1**

CEL-seq2/C1 libraries were generated as previously described ([Hashimshony et al., 2016](#)). Briefly, cells (200,000/ml), ERCC spike-ins, reagents and barcoded oligo-dT primers (Sigma-Aldrich) were loaded on a 10-17 μ m C1 Open-App microfluidic IFC (Fluidigm). Cell lysis, reverse transcription, second strand synthesis and in-vitro transcription were performed on-chip. Subsequently, harvested aRNA was pooled from 48 capture sites. After fragmentation and clean-up, 5 μ l of aRNA was used to construct final libraries by reverse transcription (SuperScript II, Thermo Fisher) and library PCR (Phusion HF, Thermo Fisher).

Drop-seq

Drop-seq experiments were performed as published (Macosko et al., 2015) and successful establishment of the method in our lab was confirmed by a species-mixing experiment (Figure S1A). For this work, J1 mES cells (100/ μ l) and barcode-beads (120/ μ l, Chem-genes) were co-flow in Drop-seq PDMS devices (Nanoshift) at rates of 4000 μ l/hr. Collected emulsions were broken by addition of perfluorooctanol (Sigma-Aldrich) and mRNA on beads was reverse transcribed (Maxima RT, Thermo Fisher). Unused primers were degraded by addition of Exonuclease I (New England Biolabs). Washed beads were counted and aliquoted for pre-amplification (2000 beads / reaction). Nextera XT libraries were constructed from 1 ng of pre-amplified cDNA with a custom P5 primer (IDT).

MARS-seq

To construct single cell libraries from polyA-tailed RNA, we applied massively parallel single-cell RNA sequencing (MARS-Seq) (Jaitin et al., 2014). Briefly, single cells were FACS-sorted into 384-well plates, containing lysis buffer and reverse-transcription (RT) primers. The RT primers contained the single cell barcodes and unique molecular identifiers (UMIs) for subsequent de-multiplexing and correction for amplification biases, respectively. Spike-in transcripts (ERCC, Ambion) were added, polyA-containing RNA was converted into cDNA as previously described and then pooled using an automated pipeline (liquid handling robotics). Subsequently, samples were linearly amplified by in vitro transcription, fragmented, and 3' ends were converted into sequencing libraries. The libraries consisted of 48 single cell pools.

SCRB-seq

RNA was stabilized by resuspending cells in RNAlater Cell Reagent (QIAGEN) and RNase inhibitors (Promega). Prior to FACS sorting, cells were diluted in PBS (Invitrogen). Single cells were sorted into 5 μ l lysis buffer consisting of a 1/500 dilution of Phusion HF buffer (New England Biolabs) and ERCC spike-ins (Ambion), spun down and frozen at -80°C . Plates were thawed and libraries prepared as described previously (Soumillon et al., 2014). Briefly, RNA was desiccated after protein digestion by Proteinase K (Ambion). RNA was reverse transcribed using barcoded oligo-dT primers (IDT) and products pooled and concentrated. Unincorporated barcode primers were digested using Exonuclease I (New England Biolabs). Pre-amplification of cDNA pools were done with the KAPA HiFi HotStart polymerase (KAPA Biosystems). Nextera XT libraries were constructed from 1 ng of pre-amplified cDNA with a custom P5 primer (IDT).

Smart-seq/C1

Smart-seq/C1 libraries were prepared on the Fluidigm C1 system using the SMARTer Ultra Low RNA Kit (Clontech) according to the manufacturer's protocol. Cells were loaded on a 10–17 μ m RNA-seq microfluidic IFC at a concentration of 200,000/ml. Capture site occupancy was surveyed using the Operetta (Perkin Elmer) automated imaging platform.

Smart-seq2

mESCs were sorted into 96-well PCR plates containing 2 μ l lysis buffer (1.9 μ l 0.2% Triton X-100; 0.1 μ l RNase inhibitor (Lucigen)) and spike-in RNAs (Ambion), spun down and frozen at -80°C . To generate Smart-seq2 libraries, priming buffer mix containing dNTPs and oligo-dT primers was added to the cell lysate and denatured at 72°C . cDNA synthesis and pre-amplification of cDNA was performed as described previously (Picelli et al., 2014b, 2013). Sequencing libraries were constructed from 2.5 ng of pre-amplified cDNA using an in-house generated Tn5 transposase (Picelli et al., 2014a). Briefly, 5 μ l cDNA was incubated with 15 μ l tagmentation mix (1 μ l of Tn5; 2 μ l 10x TAPS MgCl_2 Tagmentation buffer; 5 μ l 40% PEG8000; 7 μ l water) for 8 min at 55°C . Tn5 was inactivated and released from the DNA by the addition of 5 μ l 0.2% SDS and 5 min incubation at room temperature. Sequencing library amplification was performed using 5 μ l Nextera XT Index primers (Illumina) that had been first diluted 1:5 in water and 15 μ l PCR mix (1 μ l KAPA HiFi DNA polymerase (KAPA Biosystems); 10 μ l 5x KAPA HiFi buffer; 1.5 μ l 10mM dNTPs; 2.5 μ l water) in 10 PCR cycles. Barcoded libraries were purified and pooled at equimolar ratios.

DNA sequencing

For SCRB-seq and Drop-seq, final library pools were size-selected on 2% E-Gel Agarose EX Gels (Invitrogen) by excising a range of 300–800 bp and extracting DNA using the MinElute Kit (QIAGEN) according to the manufacturer's protocol.

Smart-seq/C1, CEL-seq2/C1, Drop-seq and SCRB-seq library pools were sequenced on an Illumina HiSeq1500. Smart-seq2 pools were sequenced on Illumina HiSeq2500 (Replicate A) and HiSeq2000 (Replicate B) platforms. MARS-seq library pools were sequenced on an Illumina HiSeq2500 using the Rapid mode. Smart-seq/C1 and Smart-seq2 libraries were sequenced 45 cycles single-end, whereas CEL-seq2/C1, Drop-seq and SCRB-seq libraries were sequenced paired-end with 15–20 cycles to decode cell barcodes and UMI from read 1 and 45 cycles into the cDNA fragment. MARS-seq libraries were paired-end sequenced with 52 cycles on read 1 into the cDNA and 15 bases for read 2 to obtain cell barcodes and UMIs. Similar sequencing qualities were confirmed by FastQC v0.10.1 (Figure S1B).

QUANTIFICATION AND STATISTICAL ANALYSIS

Basic data processing and sequence alignment

Smart-seq/C1/Smart-seq2 libraries (i5 and i7) and CELseq2/C1/Drop-seq/SCRB-seq pools (i7) were demultiplexed from the Illumina barcode reads using deML (Renaud et al., 2015). MARS-seq library pools were demultiplexed with the standard Illumina pipeline. All reads were trimmed to the same length of 45 bp by cutadapt (Martin, 2011) (v1.8.3) and mapped to the mouse genome (mm10)

including mitochondrial genome sequences and unassigned scaffolds concatenated with the ERCC spike-in reference. Alignments were calculated using STAR 2.4.0 (Dobin et al., 2013) using all default parameters.

For libraries containing UMIs, cell- and gene-wise count/UMI tables were generated using the published Drop-seq pipeline (v1.0) (Macosko et al., 2015). We discarded the last 2 bases of the Drop-seq cell and molecular barcodes to account for bead synthesis errors. For Smart-seq/C1 and Smart-seq2, features were assigned and counted using the Rsubread package (v1.20.2) (Liao et al., 2013).

Power Simulations

We developed a framework in R for statistical power evaluation of differential gene expression in single cells. For each method, we estimated the mean expression, dispersion and dropout probability per gene from the same number of cells per method. In the read count simulations, we followed the framework proposed in Polyester (Frazee et al., 2015), i.e., we retained the observed mean-variance dependency by applying a cubic smoothing spline fit to capture the heteroscedasticity observed. Furthermore, we included a local polynomial regression fit for the mean-dropout relationship. In each iteration, we simulated count measurements for the 13,361 genes for sample sizes of 2^4 , 2^5 , 2^6 , 2^7 , 2^8 and 2^9 cells per group. The read count for a gene i in a cell j is modeled as a product of a binomial and negative binomial distribution:

$$X_{ij} \sim B(p = 1 - p_0) * NB(\mu, \theta).$$

The mean expression magnitude μ was randomly drawn from the empirical distribution. 5 percent of the genes were defined as differentially expressed with an effect size drawn from the observed fold changes between microglial subpopulations in Zeisel et al. (Zeisel et al., 2015). The dispersion θ and dropout probability p_0 were predicted by above mentioned fits.

For each method and sample size, 100 RNA-seq experiments were simulated and tested for differential expression using limma (Ritchie et al., 2015) in combination with voom (Law et al., 2014) (v3.26.7). The power simulation framework was implemented in R (v3.3.0).

ERCC capture efficiency

To estimate the single molecule capture efficiency, we assume that the success or failure of detecting an ERCC is a binomial process, as described before (Marinov et al., 2014). Detections are independent from each other and are thus regarded as independent Bernoulli trials. We recorded the number of cells with nonzero and zero read or UMI counts for each ERCC per method and applied a maximum likelihood estimation to fit the probability of successful detection. The fit line was shaded with the 95% Wilson score confidence interval.

Cost efficiency calculation

We based our cost efficiency extrapolation on the power simulations starting from empirical data at different sequencing depths (250,000 reads, 500,000 reads, 1,000,000 reads; Figure 6C). We determined the number of cells required per method and depth for adequate power (80%) by an asymptotic fit to the median powers. For the calculation of sequencing cost, we assumed 5€ per million raw reads, independent of method. Although UMI-based methods need paired-end sequencing, we assumed a 50 cycle sequencing kit is sufficient for all methods. We used prices in Euro as a basis and consider an exchange course of 1:1 for the given prices in USD.

DATA AND SOFTWARE AVAILABILITY

The accession number for the raw and analyzed scRNA-seq data reported in this paper is GEO: GSE75790.

Molecular Cell, Volume 65

Supplemental Information

Comparative Analysis of Single-Cell RNA Sequencing Methods

Christoph Ziegenhain, Beate Vieth, Swati Parekh, Björn Reinius, Amy Guillaumet-Adkins, Martha Smets, Heinrich Leonhardt, Holger Heyn, Ines Hellmann, and Wolfgang Enard

Supplementary Figures

Figure S1

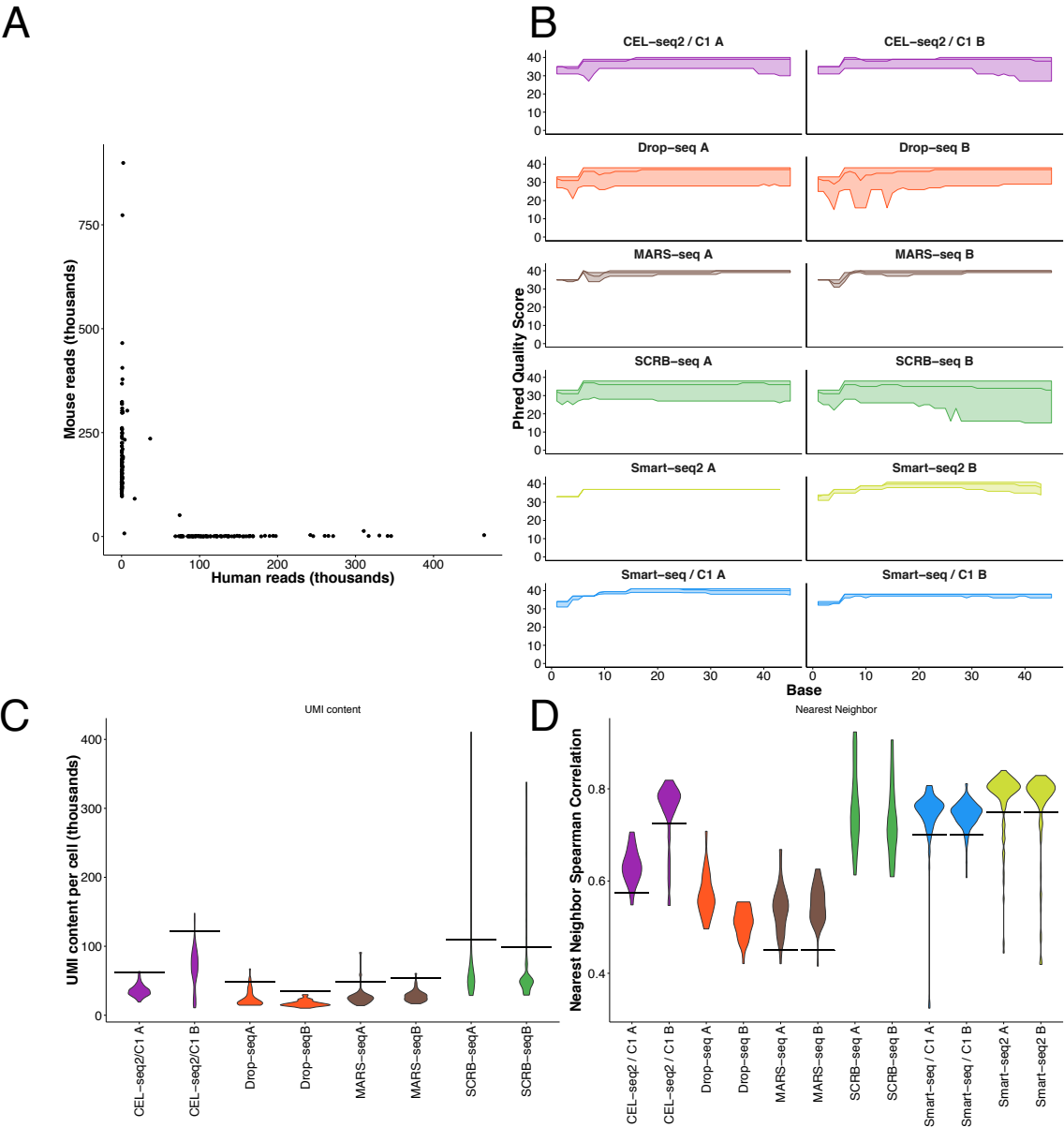


Figure S2

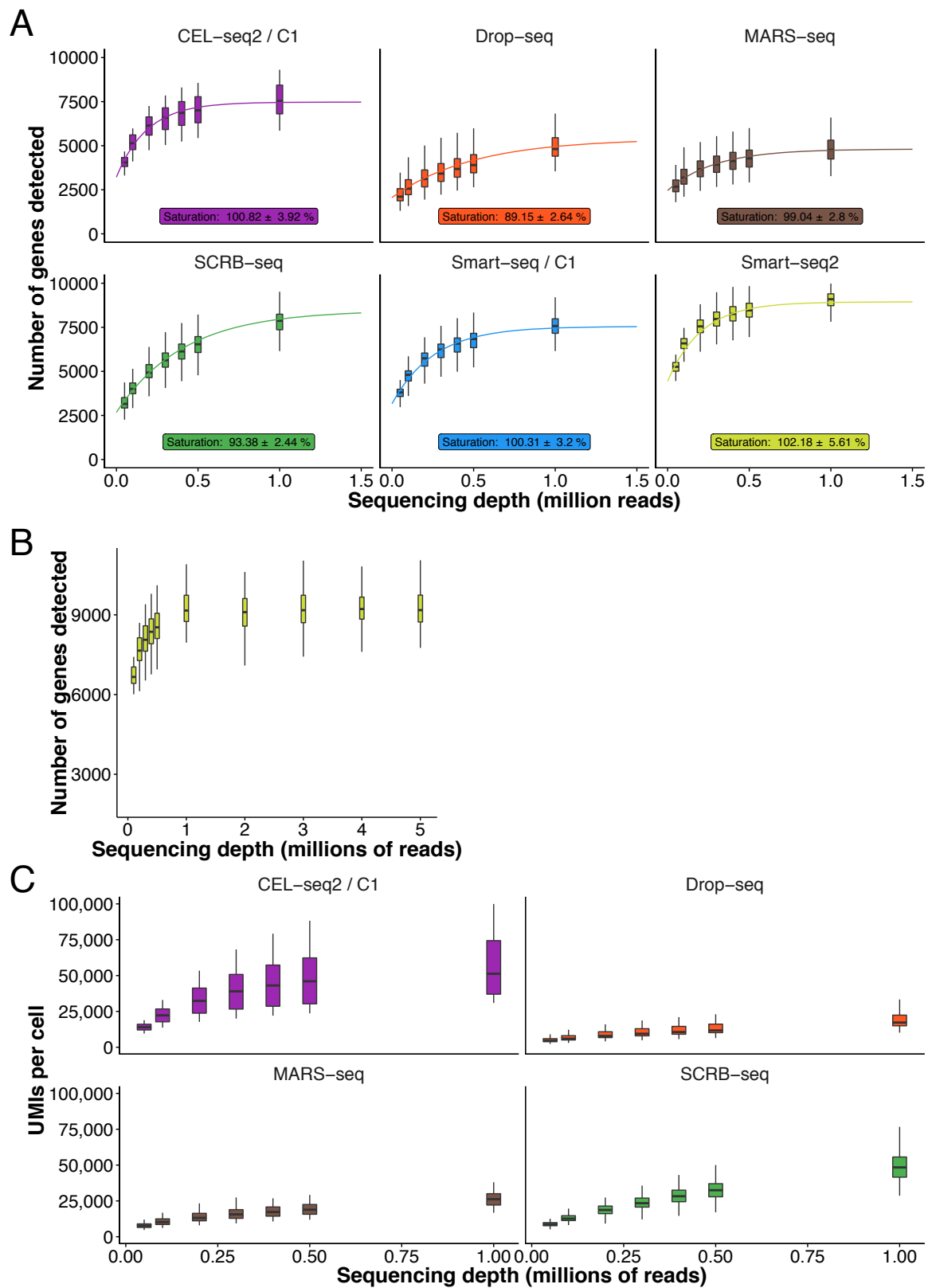


Figure S3

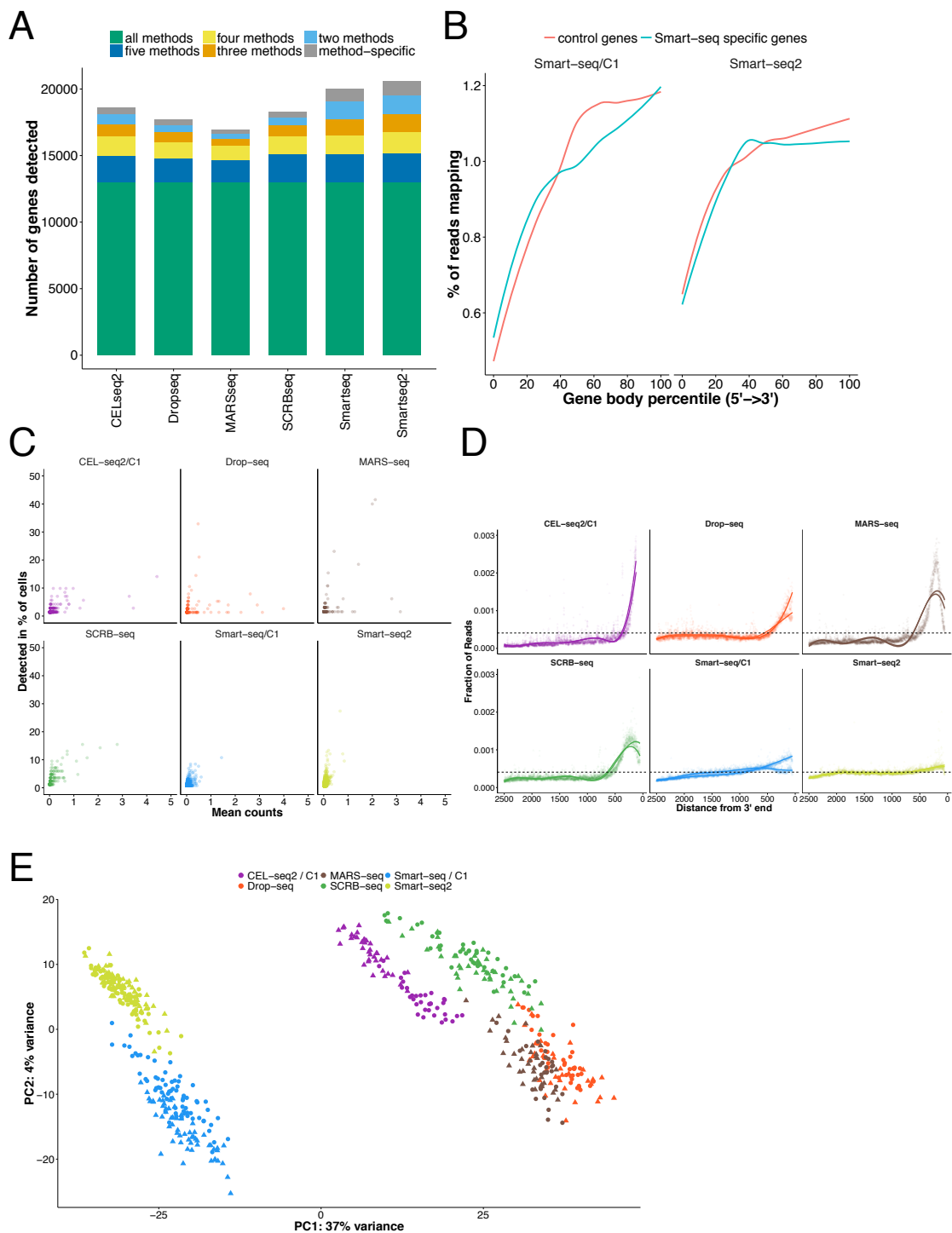


Figure S4

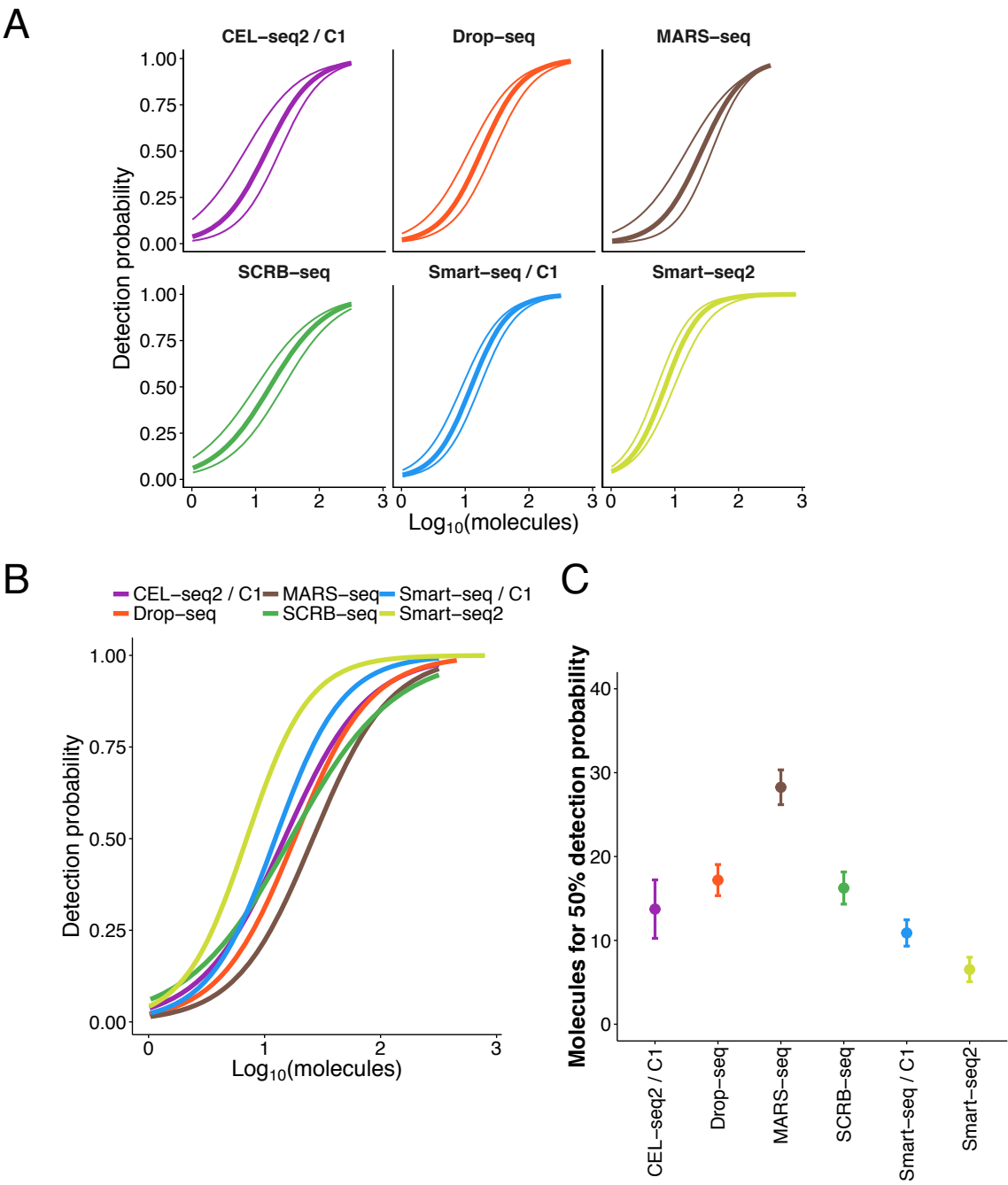


Figure S5

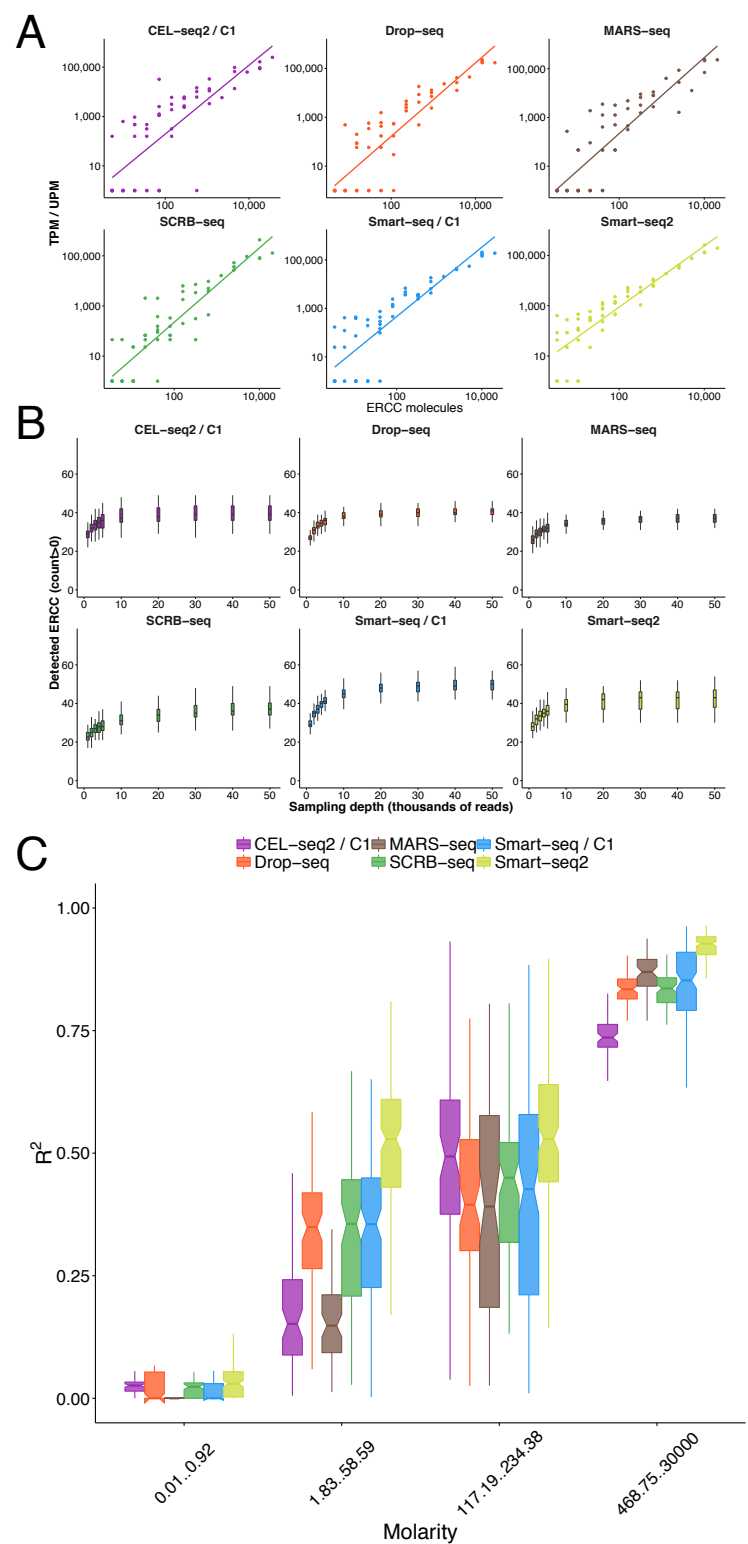


Figure S6

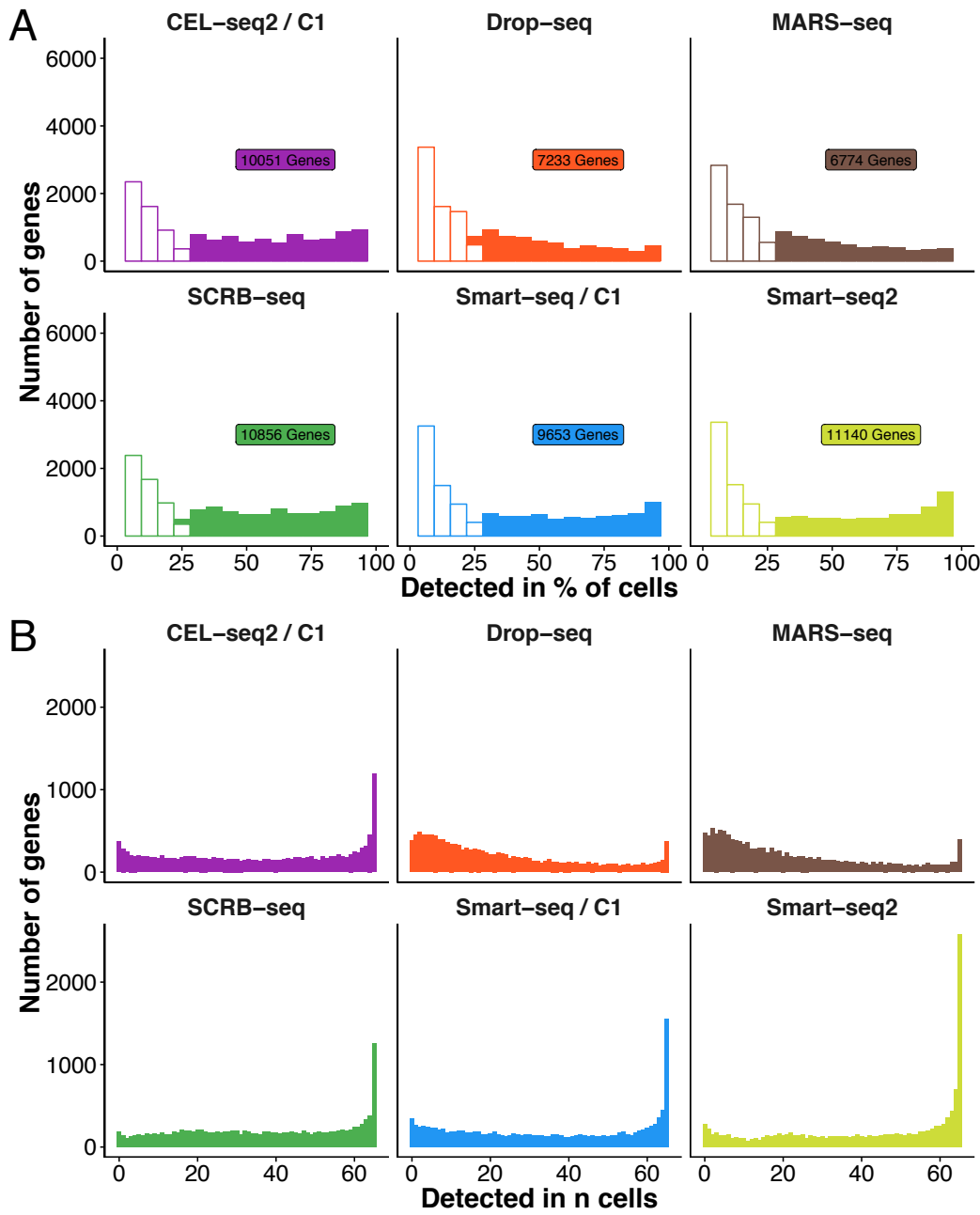


Figure S7

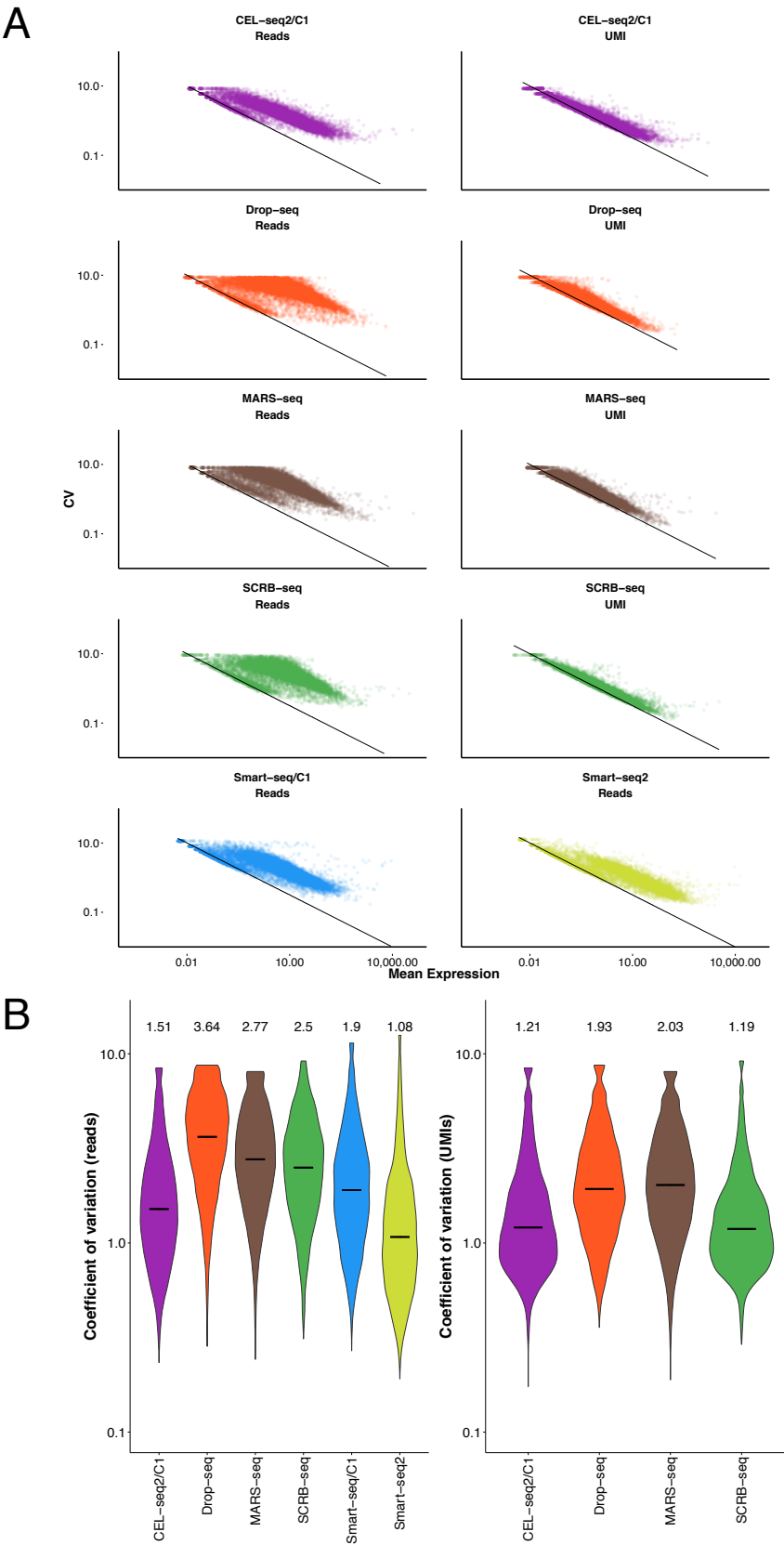


Figure S8

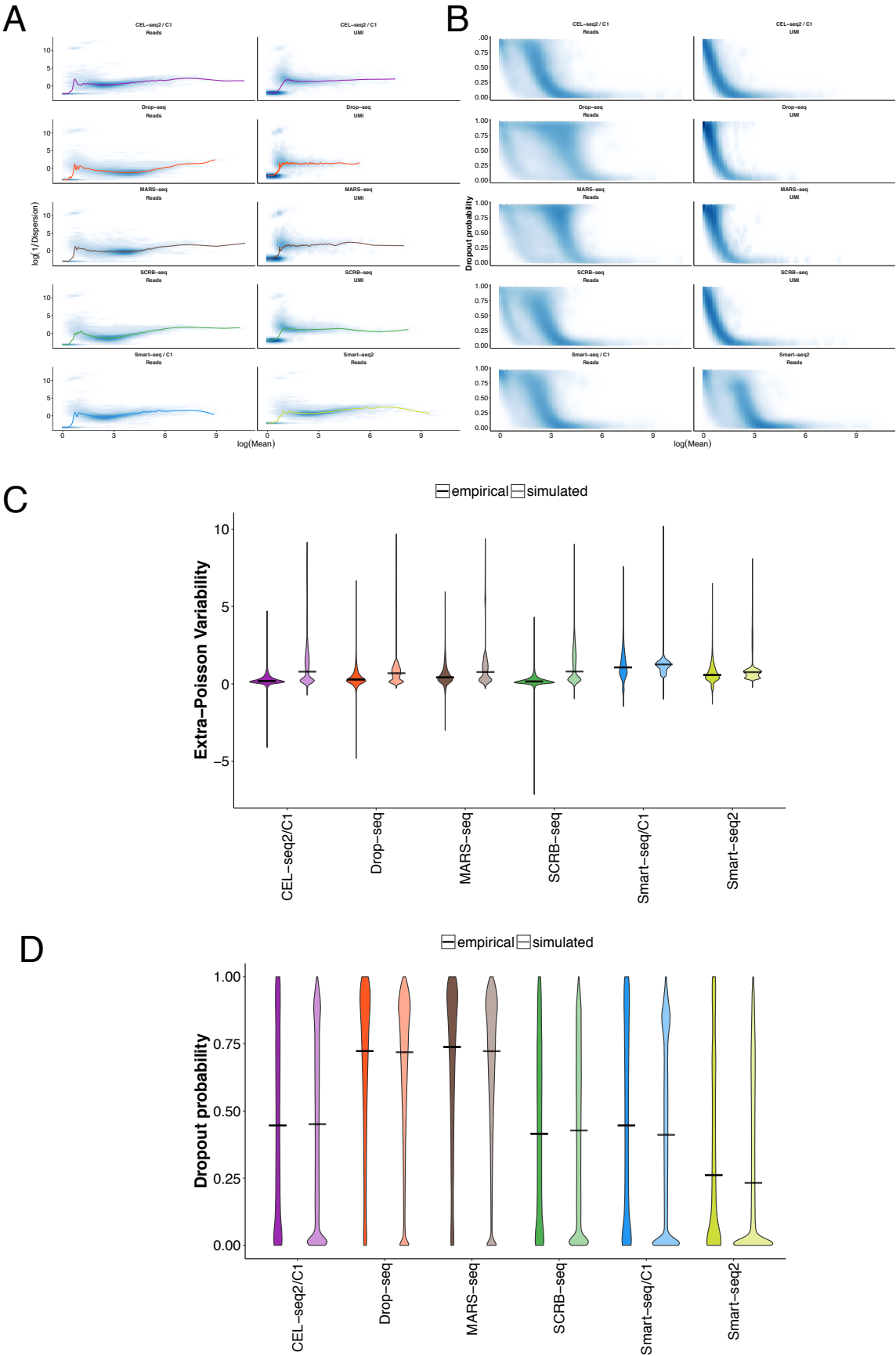


Figure S9

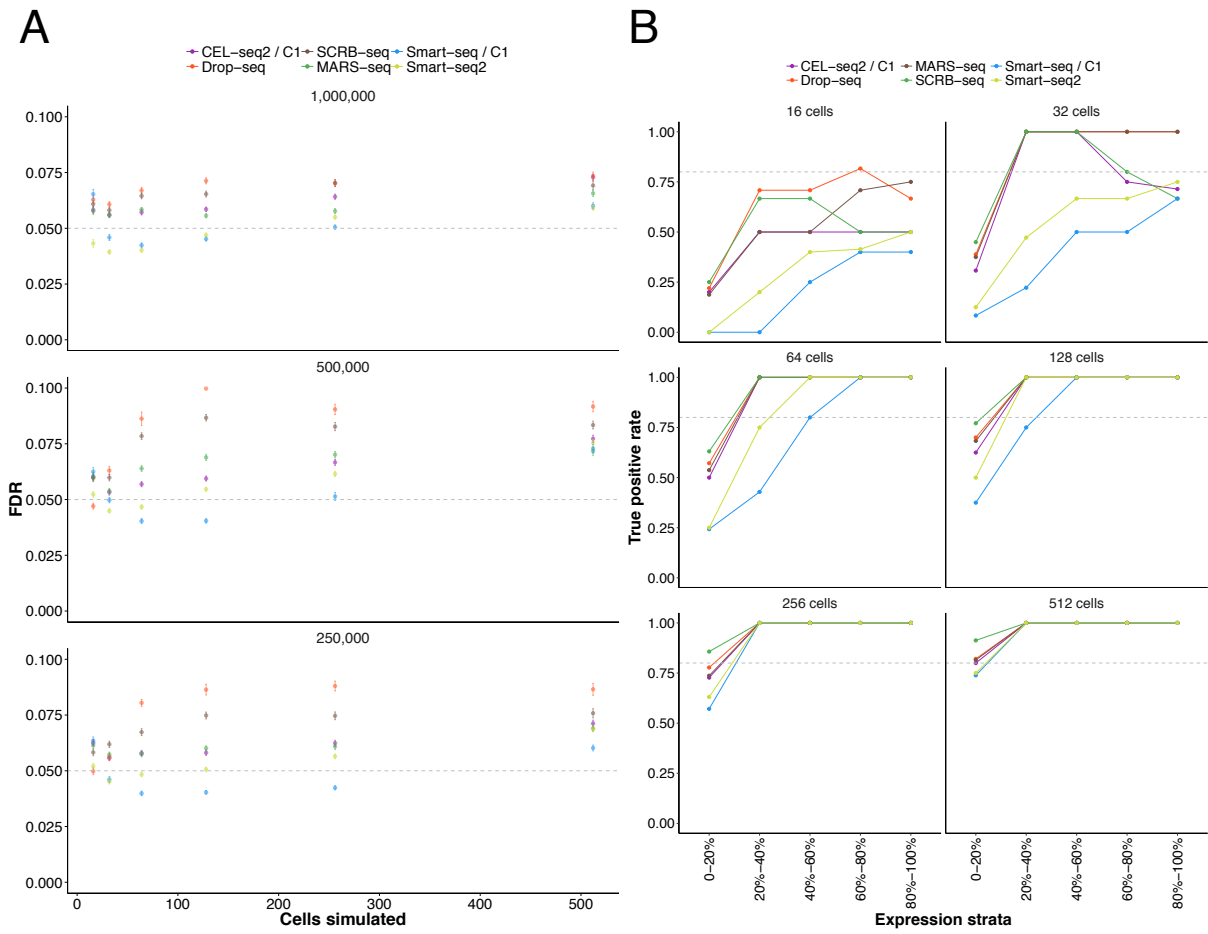
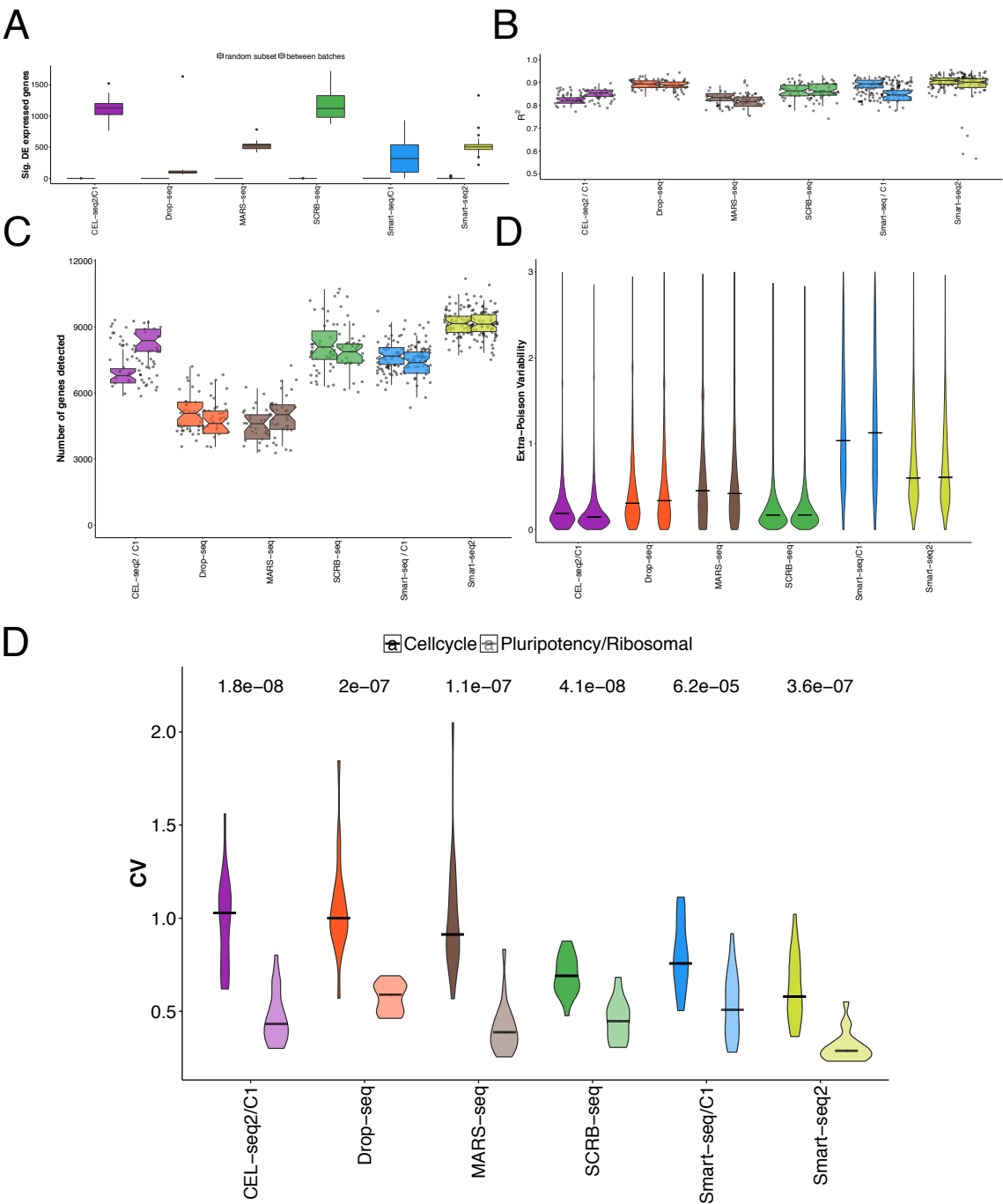


Figure S10



Supplementary Figure Legends

Figure S1 (related to Figure 1) | Quality control and filtering. **A** Drop-seq species mixing experiment using human and murine T-cells. For each cell-barcode human- and mouse read numbers are plotted. **B** Per-base quality scores were summarized using FastQC. Lines indicate median Phred quality score with upper and lower quartile shaded. **C** Total UMI content per cell, with the filter cutoff (two times mean) shown as black lines. Violin plots indicate the density of the UMI content distribution per replicate. **D** Nearest-neighbor filtering based on the maximum pairwise Spearman's rho for each cell. Violin plots indicate the density of rho distribution per replicate. Black lines indicate the employed cutoffs.

Figure S2 (related to Figure 1) | Downsampling of scRNA-seq libraries. **A** Detected genes (≥ 1 count) in relation to indicated sequencing depths. The ranges of the boxes indicate the upper and lower quartiles of cells and horizontal bars indicate the medians. **B** Boxplots of the number of detected genes in high-depth sequencing of Smart-seq2 libraries, showing a plateau above 1 million reads. **C** Boxplots of the number of detected UMIs per cell in relation to indicated sequencing depths.

Figure S3 (related to Figure 3) | Sensitivity **A** The overlap of detected genes (≥ 1 count) between methods for 65 random cells is displayed as a barplot. Colors indicate the level of overlap: Green (detected in all methods), dark blue (detected in five methods), yellow (detected in four methods), orange (detected in three methods), light blue (detected in two methods), grey (method-specific detection). **B** Gene body coverage (left to right equalling 5' to 3') of ~3000 genes detected by Smart-seq/C1 and/or Smart-seq2 (right panel) versus a random control set of 3000 genes detected by all methods. **C** Method-specific detected genes are shown as scatter plots with their rate of detection and mean counts over all cells. **D** For genes and their transcript variants of at least 2 kb length, we calculated the fraction of reads mapping to positions relative to the 3' end. For each method, we show mapping positions and a fit line per replicate. The dashed line indicates theoretical even distribution of reads across the 2.5 kb window. **(E)** Gene expression values were normalized as transcripts per million TPM or UMIs per million UPM. Principal component analysis was performed on the 1000 most variable genes to display the major variance between single cells. The 200 genes with the highest loading for PC1 were analysed and neither showed significant enrichment in GO categories (GORilla) nor in technical properties such as gene length or GC content.

Figure S4 (related to Figure 3) | Detection probabilities were estimated from ERCC dropouts, where the RNA molecule number is known. **A** Thick lines indicate the maximum-likelihood estimate of the detection probability with the thin lines showing the 95% confidence interval of the fit. **B** Shown are per-method maximum-likelihood estimates of mRNA detection probabilities. **C** Sensitivity per method estimated as the 50% probability to detect a transcript. The 95% confidence interval of estimate is displayed as error bars.

Figure S5 (related to Figure 4) | **A** Exemplary correlations of ERCC expression values (transcripts per million TPM or UMIs per million UPM) with annotated concentrations. For each method, we chose a representative cell/bead with a linear model correlation coefficient close to the median of all cells. **B** Detection of ERCC genes (≥ 1 count) in relation to sampling depth. Each boxplot represents the median, upper and lower quartile of all cells within each method. **C** Accuracy of scRNA-seq methods. ERCC expression values were correlated to their annotated molarity. Shown are the distributions of correlation coefficients (adjusted R^2 of linear regression model) across methods for bins of ERCC molarity. Each boxplot represents the median, first and third quartile for the R^2 in the indicated bin.

Figure S6 (related to Figure 5) | Gene detection sparsity. **A** For all detected genes (≥ 1 CPM) per method, we calculated the rate of detection. Histograms show this measure for detection sparsity. Filled bars represent the genes detected in at least 25% of cells of each method along with the number of these reproducibly detected genes. **B** For genes detected in at least 25% of cells of any method, we calculate the rate of detection in 65 random cells.

Figure S7 (related to Figure 5) | Variation in scRNA-seq data. **A** Gene-wise mean and coefficient of variation from all cells are shown as scatterplots for all methods. The black line indicates variance according to the poisson distribution. The two populations of genes seen for read-count data are unamplified genes (close to Poisson, one or very few reads per UMI) and amplified genes (higher CV for a given mean, several reads per UMI). **B** Gene-wise coefficient of variation (CV) of scRNA-seq data were calculated for all cells including detection dropouts. Violin plots are shown for UMI and read-count based quantification indicating the density of the distribution.

Figure S8 (related to Figure 6) | **A-B** Power simulation parameters estimated from 1 million reads per cell. **A** Mean expression and size parameters were estimated for each method and their functional relation was approximated by a smooth spline fit. **B** The dropout probability p_0 was calculated per gene and shown in relation to mean expression levels. We

fitted this relationship using a local polynomial regression. **C-D** Validation of power simulation framework. **C** Gene-wise Extra-Poisson Variability was calculated from empirical data and simulated data without addition of differentially expressed genes. Shown are the distributions with the black line indicating the median. **D** Gene-wise dropout rate distributions are shown from empirical data and simulated data. The black line indicates the median dropout rate.

Figure S9 (related to Figure 6 and Table 1) | **A** FDR. Simulations were performed using empirical mean, dispersion and dropout relationships (see Figure S8). For variable sample sizes of $n=16$, $n=32$, $n=64$, $n=128$, $n=256$ and $n=512$, we show points representing the mean FDR of 100 simulations with standard error. **B** | Stratified analysis of power. Shown are TPR for 1 million reads per cell for sample sizes $n=16$, $n=32$, $n=64$, $n=128$, $n=256$ and $n=512$ per group. Genes are grouped in five percentiles of mean expression with lines representing the median TPR of 100 simulations.

Figure S10 (related to Figure 6) | **A-D** Batch effects **A** For each method, we test for differential expression between random subsets of 25 cells per group (left box) and subsets of 25 cells of each batch (right box) in 20 permutations using limma. Shown are the number of significantly differentially expressed genes ($FDR < 0.01$) as boxplots. **B** Sensitivity is shown as the number of detected genes (≥ 1 count) per batch. **C** Accuracy is shown per batch as the correlation coefficient of observed expression (TPM/UPM) to annotated ERCC molecule numbers. **D** Precision is shown per batch as the Extra-Poisson Variability for the common 13,361 genes. For 3' counting methods, UMI quantification is shown. The distribution was only shown between values of 0 and 3 to make differences more visible. **D** Cell cycle analysis. For each method, we show the coefficient of variation (CV) for a set of 19 cell cycle genes previously found to be variable in 2i/LIF cultured mESCs (Kolodziejczyk, 2015) (left violin) compared to 19 ribosomal and pluripotency genes. Numbers above the violins indicate p-values of a t-test between the two groups.

Supplementary Tables

Method	CEL-seq2/C1	Drop-seq	MARS-seq	SCRB-seq	Smart-seq/C1	Smart-seq2
Single-cell isolation	automated in the C1 system	droplets	FACS	FACS	automated in the C1 system	FACS
ERCC spike-ins	yes	no	yes	yes	yes	yes
UMI	6 bp	8 bp	8 bp	10 bp	no	no
Full-length coverage	no	no	no	no	yes	yes
1st strand synthesis	oligo-dT	oligo-dT	oligo-dT	oligo-dT	oligo-dT	oligo-dT
2nd strand synthesis	RNAseH / DNA Pol	template switching	RNAseH / DNA Pol	template switching	template switching	template switching
Amplification	IVT	PCR	IVT	PCR	PCR	PCR
Imaging of cells possible	yes	no	no	no	yes	no
Protocol usable for bulk	yes	no	yes	yes	yes	yes
Sequencing	paired-end	paired-end	paired-end	paired-end	single-end	single-end
Library cost /cell	~9.5€	~0.1€	~1.3€	~2€	~25€	~3/30*

Table S1 (related to Figure 2): Overview of single-cell RNA-seq methods.

* in-house produced Tn5 / commercial Tn5

Results

2.5 Ubiquitome analysis reveals PAF15 as a specific ubiquitination target of UHRF1 in embryonic stem cells

Ubiquitome analysis reveals PAF15 as a specific ubiquitination target of UHRF1 in embryonic stem cells

Martha Smets^{1#}, Elisabeth Schmidtmann^{1#}, Ignasi Forné², Weihua Qin¹, Christopher B Mulholland¹, Joel Ryan¹, Axel Imhof² and Heinrich Leonhardt^{1*}

¹Department of Biology II and Center for Integrated Protein Science Munich (CIPSM), Ludwig-Maximilians-Universität München, Großhaderner Str. 2, 82152 Planegg-Martinsried, Germany

²BioMedical Center (BMC), Department of Molecular Biology, Ludwig-Maximilians-Universität München, Großhaderner Str. 9, 82152 Planegg-Martinsried, Germany

Co-first authors.

*Correspondence: Tel: +49 89 2180 74232; Fax: +49 89 2180 74236, h.leonhardt@lmu.de (H.L.)

Running title: UHRF1 ubiquitinates PAF15

Keywords: Ubiquitome, PAF15, UHRF, KIAA0101, S-phase, TLS

Abstract

Ubiquitination is a pivotal posttranslational modification of proteins associated with distinct physiological effects such as functional modulation, subcellular localization and targeted degradation. The E3 ligase UHRF1 is an essential epigenetic cofactor for DNMT1 dependent maintenance DNA methylation, which provides a binding platform for DNMT1 by both cooperative binding of histones and hemi-methylated DNA as well as by ubiquitinating histone H3. However, the catalytic role of UHRF1 in non-epigenetic pathways has been discovered only recently. Here, we conduct a comprehensive screen in mouse embryonic stem cells to identify novel ubiquitination targets of UHRF1 and its paralogue UHRF2, which are highly similar in both sequence and structure. We find differentially ubiquitinated peptides involved in a variety of biological processes such as transcriptional regulation and DNA damage response. Most importantly, we discover PCNA associated factor PAF15 (mouse: Pclaf, Ns5atp9, human: KIAA0101, OEATC-1) as a specific ubiquitination target of UHRF1. Although the function of PAF15 ubiquitination in translesion DNA synthesis (TLS) is well characterized, the respective E3 ligase remained unknown until now. We show that UHRF1 ubiquitinates PAF15 at Lys 15 and Lys 24 and further promotes its binding to PCNA during late S-phase. In summary, we uncover novel UHRF1-dependent ubiquitination targets thereby shedding light on its functional relevance beyond epigenetic regulation.

Introduction

Posttranslational modifications such as ubiquitination greatly affect protein function in a variety of cellular processes. The reversible conjugation of ubiquitin molecules to a target protein has distinct physiological effects such as destabilization of target proteins, altered protein trafficking and functional modulation (Hicke and Dunn 2003; Pickart 2004; Sun and Chen 2004; Pickart and Fushman 2004). Ubiquitination of lysine residues is mediated in a E1–E2–E3 tri-enzyme cascade, where ubiquitin transfer from a E2~Ub intermediate to a lysine on a substrate is mediated by E3 ligase enzymes. E3 ligase activity is often endowed in a Really Interesting New Gene (RING) domain (Brown et al. 2015), which is present in Ubiquitin-like PHD and RING finger domain-containing protein 1 (UHRF1) and its paralogue UHRF2. UHRF1 (also known as NP95 or ICBP90) is not only a well-characterized factor in DNA methylation maintenance, rendering it essential for early embryonic development, but also for cell cycle regulation and genome stability (Muto 2002; Jenkins et al. 2005).

First, UHRF1 targets maintenance DNA methyltransferase 1 (DNMT1) to newly synthesized DNA in heterochromatin after replication (Bostick et al. 2007; Sharif et al. 2007; Achour et al. 2008), by cooperative binding of repressive H3K9me3 marks and hemimethylated DNA (Liu et al. 2013) and by ubiquitination of H3 tails on K18 (K23 in *Xenopus*), which is specifically recognized (and bound) by the ubiquitin interacting motif (UIM) in the TS domain of DNMT1 (Nishiyama et al. 2013; Qin et al. 2015).

Second, UHRF1 plays a role in cell cycle progression as shown by its co-localization with proliferating cell nuclear antigen (PCNA) during S phase (Uemura et al. 2000) and the increased sensitivity of UHRF1-deficient embryonic stem cells (ESCs) towards treatment with the replication-inhibiting reagent hydroxyurea (Muto 2002).

Finally, UHRF1 has a critical role in maintenance of genome stability (Muto 2002; Luo et al. 2013) by recognizing and binding DNA interstrand crosslinks (ICLs) and thereby inducing repair pathways such as the Fanconi anemia pathway (Liang et al. 2015; Tian et al. 2015). Further, UHRF1 is important for the repair of DNA double strand breaks in a cell cycle dependent manner (Zhang et al. 2016).

Although numerous reported functions of UHRF1 involve ubiquitination activity of target proteins, such as DNMT1 (Qin, Leonhardt, and Spada 2011a; Du et al. 2010) and histone H3 (Nishiyama et al. 2013;

Qin et al. 2015; Citterio et al. 2004), no comprehensive screen of ubiquitination targets of UHRF1 was published so far.

Here, we screen for specific ubiquitination targets of UHRF1 by comparing the ubiquitome of wildtype (wt), UHRF1- and UHRF2-deficient mouse ESCs. With an antibody-dependent enrichment of ubiquitin remnant motif-containing peptides followed by isobaric-labeling based quantitative mass spectrometry, we find both known and novel E3 ligase substrates of UHRF1 involved in a variety of biological processes such as RNA processing, DNA methylation and DNA damage repair. Our results uncover that PCNA-interacting factor (PAF15) 15 (Yu et al. 2001) is a ubiquitination target of UHRF1 but not UHRF2. Ubiquitination of PAF15 is well characterized to be important in replication block bypass by regulating the recruitment of translesion DNA synthesis (TLS) polymerases (Povlsen et al. 2012) but the respective E3-ligase was not identified until now. We demonstrate that UHRF1-dependent ubiquitination promotes binding of PAF15 to PCNA, thereby unraveling a novel function of UHRF1 in regulating DNA damage response.

Results

Ubiquitome of mouse embryonic stem cells deficient for UHRF1

To identify specific ubiquitination targets of UHRF1 in ESCs, we compared the ubiquitome of UHRF1- and UHRF2-deficient cells relative to wt. Enrichment of formally ubiquitinated tryptic peptides was performed with a specific K-gly-gly antibody, which recognizes a remnant gly-gly motif on the formerly ubiquitinated lysine residue (Xu, Paige, and Jaffrey 2010). For relative peptide quantification in mass spectrometry, enriched peptide fractions were labeled with isobaric tandem mass tag (TMT) reagents and pooled for subsequent LC-MS/MS analysis (Figure 1A). In total, we quantified 1248 K-gly-gly-containing peptides across two measurements (_A, _B, Supplementary Table S1). 53 peptides show high abundance differences with an intensity change of 3 ($\log_2 = 1.58$) or higher in at least two replicates (Figure 1B). We detect both enriched and de-enriched ubiquitinated peptides in UHRF1-depleted cells compared to wt. The abundance of K-gly-gly peptides is not necessarily reflected by altered protein expression (Figure 1C), thus the observed differences are due to posttranslational effects.

PAF15 as a ubiquitination target of UHRF1

For statistical analysis of UHRF1 ubiquitination targets, we compared peptides quantified across all measured samples and found differentially ubiquitinated peptides in both *Uhrf1*^{-/-} and *Uhrf2*^{-/-} cells (Supplementary Table S2). Peptides with significant ubiquitination changes can be assigned to GO terms such as transcriptional regulation, cell cycle regulation and DNA damage response (Figure 2A and 2B), indicating that the UHRF family is involved in the regulation of a variety of different proteins. We found 94 differentially regulated peptides in *Uhrf1*^{-/-} cells of which 62,8% is not found to be differentially ubiquitinated in the *Uhrf2*^{-/-} (Supplementary Figure S1). Notably, in *Uhrf1*^{-/-} cells the highest loss of ubiquitination was observed for lysine 15 and 24 of PAF15 (Figure 3A), whereas the ubiquitination state of PAF15 in *Uhrf2*^{-/-} cells remained unchanged (Figure 2C, D), indicating that PAF15 is a ubiquitination target of UHRF1.

The RING domain of UHRF1 ubiquitinates PAF15 on Lys 15 and Lys 24

We confirmed UHRF1 as the E3-ligase of PAF15 by analysis of a *Uhrf1*^{-/-} ESC line with a different genetic background (E14). Due to its low protein abundance, we performed immunoprecipitation experiments to enrich PAF15. In wt ESCs, PAF15 is mono- and mainly di-ubiquitinated, whereas in *Uhrf1*^{-/-} cells PAF15 is unmodified. Ubiquitination of PAF15 is reestablished upon expression/reintroduction of wt UHRF1-GFP but not UHRF1-GFP H730A, a mutation with reduced E3 ligase activity (Qin et al. 2015) (Figure 3B). Thus, UHRF1 does not recruit the E3 ligase but rather directly modifies PAF15 with/by/dependent on its RING domain.

Endogenous PAF15 localization throughout S-phase

PAF15 was originally found to be associated with PCNA in a yeast-two-hybrid screen (Yu et al. 2001), while UHRF1 is mainly associated with replicating heterochromatin (Uemura et al. 2000; Papait et al. 2007). As the interaction with PCNA is essential for PAF15 ubiquitination (Povlsen et al. 2012), we investigated the spatial distribution of UHRF1 and PAF15 at sites of replication. With super-resolution microscopy, we showed that PAF15 and PCNA co-localize predominantly in late S-phase in C2C12 myoblasts (median correlation coefficient = 0,55; Figure 4A, 4C). Likewise, PAF15 and UHRF1 also display the closest proximity in late S-phase (median correlation coefficient = 0,35; Fig. 4B, D, Supplementary Material, SFig. 2). Taken together, UHRF1 co-localizes with PAF15 at sites of PCNA foci in late S-phase, where heterochromatic regions are replicated and thereby could ubiquitinate PAF15 in a cell cycle dependent manner.

PAF15-PCNA interaction is promoted by UHRF1 dependent ubiquitination

To investigate the role of UHRF1 for PAF15 localization, we performed immunofluorescence stainings and found PAF15 co-localizing with PCNA in wt and *UHRF2* depleted ESCs, whereas in *Uhrf1*^{-/-} ESCs, PAF15 displays a diffuse pattern in late S-phase (Fig. 5A). Furthermore, the subcellular localization of PAF15 in *Uhrf1*^{-/-} is restored by expressing UHRF1-GFP wt (Fig. 5B).

To test if PAF15 binding to PCNA is promoted by mono-ubiquitination on positions Lys 15 and Lys 24, we performed a rescue experiment in *PAF15*^{-/-} ESCs with GFP-PAF15 wt and double-mutant GFP-PAF15 K15R.K24R (dm). Interestingly, GFP-PAF15 wt co-localizes with PCNA, whereas GFP-PAF15 dm is diffusely distributed in the nucleus and only to a little extent associated with PCNA in late S-phase (Supplementary Fig. S3A). Consistent results were obtained in a fluorescence-three-hybrid (F3H) assay (Herce et al. 2013), where RFP-PCNA is recruited to GFP-PAF15 wt, but not to GFP-PAF15 dm (Supplementary Fig. S3B) confirming that the ubiquitination mark promotes PAF15-PCNA interaction.

Discussion

E3 ligase proteins mediate the final step of ubiquitin attachment to a target protein, thereby influencing protein degradation, cell cycle progression, DNA repair and transcription (Hicke and Dunn 2003; Pickart 2004; Sun and Chen 2004; Pickart and Fushman 2004).

In this study, we investigated specific ubiquitination targets of E3 ligase UHRF1 in mouse embryonic stem cells. We used a proteomics approach to perform an unbiased, proteome-wide and site-specific analysis of ubiquitination changes (Udeshi et al. 2013). Since the paralogue UHRF2 is highly similar to UHRF1 in both sequence and structure, we compared the ubiquitome of *Uhrf1* and *Uhrf2* knock-out cells to exclude redundancy.

We find numerous differentially ubiquitinated proteins that encompass biological processes such as transcriptional regulation, RNA binding, DNA damage response and cell cycle regulation. We find ubiquitination targets of both UHRF1 and UHRF2 such as HSP90, DNMT3b (Ding et al. 2016; Meilinger et al. 2009; Quenneville et al. 2011) as well as UHRF1 specific targets such as UHRF1 itself, Trim28 and H3K18 (Qin et al. 2015; Quenneville et al. 2011; Citterio et al. 2004). (Further, we find differentially ubiquitinated histones: H3, H2B, H2A, not different: H1, which is consistent with studies of UHRF1 in vitro and in vivo ubiquitination (Harrison et al. 2016).)

Most importantly, we find PAF15 as a protein undergoing highest loss of ubiquitination upon UHRF1 depletion. Mono-ubiquitination of PAF15 at Lys 15 and 24 has been associated with TLS inhibition by masking TLS polymerase binding sites on PCNA during undisturbed S-phase (Povlsen et al. 2012). Stalled replication caused by DNA lesions leads to PAF15 ubiquitin chain elongation and subsequent degradation, which is the basis for TLS polymerase recruitment to PCNA (Povlsen et al. 2012).

However, the E3 ligase responsible for PAF15 mono-ubiquitination remained unknown until now (Xie, Yao, and Dong 2014). Here, we show that the RING domain of UHRF1 ubiquitinates PAF15 at Lys 15 and 24 and influences its association with PCNA throughout S-phase.

The PIP domain dependent PCNA interaction of PAF15 is necessary for its ubiquitination (Povlsen et al. 2012) and our high resolution microscopy analyses revealed PAF15 co-localization with PCNA and UHRF1 exclusively during late S-phase. Thus, we suggest that the ubiquitination takes place in a cell cycle dependent manner.

Furthermore, both UHRF1 depletion and mutation of the lysine residues result in loss of PAF15 association with PCNA, which hints towards a role for PAF15 ubiquitination in stabilizing the PAF15-PCNA complex during replication. We speculate that the ubiquitination mark could potentially be recognized by adjacent protein domains such as the ubiquitin interacting motif (UIM) of DNMT1 (Qin et al. 2015).

In summary, this study provides a novel aspect of UHRF1 in regulating replication block bypass by ubiquitinating PAF15, thereby uncovering an additional mode-of-action of UHRF1 in replication-dependent DNA damage response.

Experimental Procedures

Cell culture and transfection

Mouse J1 and E14 ESCs were cultured without feeder cells in gelatinized flasks as described before (Meilinger et al. 2009). Culture medium was either supplemented with 1000 U/ml recombinant leukemia inhibitory factor LIF (Millipore) or additionally with 1 μ M MEK inhibitor PD0325901, 3 μ M GSK-3 inhibitor CHIR99021 (2i, Axon Medchem) to keep ESCs in unprimed state. E14 ESCs and E14 *Uhrf1* knockout cells stably rescued with either UHRF1-GFP (wt) or RING domain point mutant UHRF1-GFP H730A were described previously (Qin et al. 2015).

Somatic cell lines used in this study were BHK cells containing multiple lac operator repeats (Tsukamoto et al. 2000) and C2C12 mouse myoblast cells (Yaffe and Saxel 1977). All cell lines were grown in a humidified atmosphere at 37°C and 5% CO₂, in Dulbecco's modified Eagle's medium (DMEM) supplemented with 1 μ M gentamycin and 10% (BHK) or 20% (C2C12) fetal calf serum. All cell lines were tested for mycoplasma on a regular basis.

ESCs were transfected with Lipofectamine® 3000 reagent (Thermo Fisher Scientific) according to the manufacturer's instructions. BHK cells were transfected using polyethylenimine (Sigma) according to the manufacturer's instructions.

Generation of *Uhrf1*^{-/-}, *Uhrf2*^{-/-} and *PAF15*^{-/-} ESC lines

To generate *PAF15*, *Uhrf1* and *Uhrf2* knock-out ESC lines (J1), we used the MIN tag strategy (Mulholland et al. 2015). In brief, we used a genome engineering strategy based on a CRISPR/Cas assisted in-frame insertion of an *attP* site, which we refer to as the multifunctional integrase (MIN) tag. At the genetic level, the MIN-tag serves as an attachment site for the serine integrase Bxb1 that can be used to recombine a knockout cassette into the genomic locus.

Mammalian expression constructs

Fusion constructs were generated using enhanced green fluorescent protein (GFP) or monomeric red monomeric cherry (Ch). The PAF15 wt sequence was amplified from E14 cDNA. GFP-PAF15 K15R.K24R double mutant (dm) expression construct was derived from the corresponding wt constructs by overlap extension PCR (Ho et al. 1989). Other constructs used in this study were UHRF1-GFP (Qin, Leonhardt, and Spada 2011b), RFP-PCNA (Easwaran et al. 2004) and pGBP-LacI (Herce et al. 2013).

Protein extraction and sample preparation for mass spectrometry

J1 wt, *Uhrf1*^{-/-} and *Uhrf2*^{-/-} mouse embryonic stem cells were cultured under serum/LIF conditions. For whole cell proteome analysis, 10⁶ cells were harvested in biological quadruplicates and further processed using the iST Sample Preparation Kit (PreOmics).

Enrichment of K-gly-gly peptides

Proteins were extracted from 2 × 10⁷ cells per sample and digested to peptides resulting in a K-gly-gly motif at former sites of ubiquitination, which was then used for antibody dependent enrichment as described in (Udeshi et al. 2013). In brief, cells were lysed in urea lysis buffer (8 M urea, 50 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM EDTA, 1 × Protease inhibitor, 50 μM PR-619, 1 mM chloroacetamide, 1 mM PMSF) and protein concentration was determined using a 660 nm Protein Assay (Pierce™). Proteins were reduced using 5 mM DTT, alkylated with 10 mM chloroacetamide and digested overnight using Lys-C (Wako Chemicals, 1:250 enzyme/protein ratio) and Trypsin (TPCK-treated, Worthington Biochem, 1:50 enzyme/protein ratio). Peptides were desalted using 200 mg tC18 Sep Pak Cartridges (Waters) and eluates were dried completely by vacuum centrifugation. For enrichment of K-gly-gly peptides, peptides were reconstituted in IAP buffer (50 mM MOPS pH 7.2, 10 mM sodium phosphate, 50 mM NaCl) and incubated for 1 h at 4°C with 120 μg α-K-gly-gly antibody (Cell Signaling Technology) crosslinked to protein G sepharose beads (Roche) with dimethyl pimelimidate dihydrochloride (DMP, Sigma). Beads were washed twice with IAP buffer and twice with phosphate buffered saline (Sigma) and peptides were eluted in 0.15% trifluoroacetic acid (TFA). Enriched peptide fractions were labeled using isobaric Tandem Mass Tag™ (TMTsixplex™, Thermo Fisher Scientific) reagents according to the manufacturer's instructions and pooled into one sample. Subsequently, the sample complexity was reduced by high pH reversed-phase chromatography (High pH Reversed-Phase Peptide Fractionation Kit, Pierce™). Peptides were separated to five fractions based on their hydrophobicity with buffers containing 17.5%, 20%, 22.5%, 25% or 30% acetonitrile in 0.1% triethylamine, respectively.

Liquid Chromatography Coupled to Tandem Mass Spectrometry

For mass spectrometry analysis, desalted peptide fractions were injected in an Ultimate 3000 RSLCnano system (Thermo) and separated in a 15-cm analytical column (75 μm ID packed in-house with ReproSil-Pur C18-AQ 2.4 μm from Dr. Maisch) with a 60 min gradient from 5 to 40% acetonitrile in 0.1% formic acid. The effluent from the HPLC was directly electrosprayed into a Qexactive HF (Thermo) operated in data dependent mode to automatically switch between full scan MS and MS/MS acquisition. Survey full scan MS spectra (from m/z 350–1400) were acquired with resolution R=120,000 at m/z 400 (AGC target of 3 × 10⁶). The 10 most intense peptide ions with charge states between 3 and 6 were sequentially isolated (window 0.7 m/z) to a target value of 1 × 10⁵, with resolution R=30,000, fragmented at 32% normalized collision energy and fixed first mass 100 m/z.

Typical mass spectrometric conditions were: spray voltage, 1.5 kV; no sheath and auxiliary gas flow; heated capillary temperature, 250°C; ion selection threshold, 33.000 counts.

Computational data analysis

Raw data analysis was performed using the MaxQuant software suite version 1.5.2.8 (Cox and Mann 2008). Peptide sequences were searched against the UniprotKB mouse proteome database (Swissprot) (UniProt Consortium 2015). Trypsin/P and Lys-C derived peptides with a maximum of three missed cleavages and a protein false discovery rate of 1% were set as analysis parameters. Carbamidomethylation of cysteine residues was considered a fixed modification, while oxidation of methionine, protein N-terminal acetylation and Gly-Gly modification of lysines were defined as variable modifications. For whole cell extract analysis, peptide/protein intensities were quantified based on MS1 intensities with the MaxLFQ algorithm (Cox et al. 2014). Reporter ions derived from the fragmented tandem mass tag were quantified on MS2 level with a minimum precursor intensity fraction of 75% and a reporter mass tolerance of 0.01 Da. Lot-specific reporter ion isotopic distributions of the TMT label reagents were used as isotopic correction factor.

Quantified K-gly-gly peptides were further evaluated using R (R Core Team 2016) and Perseus version 1.5.4.1 or 1.5.5.1 (Tyanova et al. 2016). The dataset was filtered for common contaminants classified by the MaxQuant algorithm and only proteins quantified across both biological replicates were subjected to statistical analysis. Differentially ubiquitinated peptides were identified using the Limma software package (Ritchie et al. 2015; Smyth 2004) after variance stabilization normalization (vsd) of peptides intensities (Huber et al. 2002).

Co-immunoprecipitation and Western Blot

For Western Blot analysis, 10⁷ ESCs cultured in serum/LIF conditions were lysed in standard lysis buffer (20 mM Tris pH 7.5, 150 mM NaCl, 0.5 mM EDTA, 2 mM MgCl₂, 0.5% NP40, 2 mM PMSF) supplemented with 1x Protease inhibitor, 1 U/μl benzonase, 50 μM PR-619 and 2.5 mM NEM. PAF15 was enriched from whole cell lysate using an anti-PAF15 antibody (Santa Cruz, sc-390515). Enriched proteins were separated on a SDS-PAGE (15% PAA) and transferred to a PVDF membrane (Millipore). PAF15 was detected using anti-PAF15 (1:500), a horseradish peroxidase conjugated anti-mouse secondary antibody (Dianova, 1:5,000) and Pierce ECL substrate (Fisher Scientific).

Immunofluorescence staining and confocal microscopy

Immunostaining was performed as described previously (Solovei, Irina, and Marion 2010). Cells cultured on coverslips were fixed with 4% paraformaldehyde for 10 min, washed with PBS-T (PBS, 0.02% Tween20) and permeabilized with 100% methanol. Both primary and secondary antibodies were diluted in blocking solution (PBS-T, 2% BSA). Coverslips with cells were incubated with primary

and secondary antibody solutions in dark humid chambers for 1 h at RT; washing steps after primary and secondary antibodies were done with PBS-T. For DNA counterstaining, coverslips were incubated in a solution of DAPI (1 µg/ml) in PBS. Coverslips were mounted in antifade medium (Vectashield, Vector Laboratories) and sealed with nail polish. For immunolabeling, the following primary antibodies were used: anti-PCNA (Rottach et al. 2008), anti-PAF15 (Santa Cruz, sc-390515) and anti-UHRF1 (Citterio et al. 2004). Secondary antibodies were anti-mouse conjugated to fluorophore 594 (Invitrogen), anti-rat conjugated to Alexa647 (Invitrogen). Single optical sections were collected using a Leica TCS SP5 confocal microscope equipped with Plan Apo 63×/1.4 NA oil immersion objective and lasers with excitation lines 405, 488, 561 and 633 nm.

Super-resolution microscopy

Cells were initially found and staged in S-phase based on their distribution of PCNA signal on a DeltaVision Elite system, equipped a 62x/1.42 PlanApo objective an interline CCD camera. To perform super-resolution structured illumination microscopy, stage coordinates of selected cells were then transferred to a DeltaVision OMX V3 3D-SIM system (Applied Precision Imaging, GE Healthcare), equipped with a 100x/1.40 NA PlanApo oil objective, three Cascade II EMCCD cameras (Photometrics), and 405-, 488-, and 594-nm laser lines. Structured Illumination (SI) images stacks consisting of 15 images per plane (five phases, at three different angles) were acquired with a z-step size of 125 nm. SI raw data were reconstructed and deconvolved with the SoftWorX 4.0 software package (Applied Precision). Registration of the three different channels was performed with the Multiview Reconstruction plugin in Fiji, using images of the nuclear pore complex stained with CF405-, Alexa488-, and Alexa594-conjugated secondary antibodies. Registered images were manually cropped to include one cell per image, background subtracted, scaled to 8-bit based on minimum and maximum pixel intensities, and colocalization analysis was performed in Fiji using the Coloc2 plugin on 5 central slices of the image stacks, taking the Pearson's Correlation coefficient without threshold as a readout of colocalization.

ACKNOWLEDGEMENTS

This work was supported by grants from the Deutsche Forschungsgemeinschaft [grant number SFB 646/B10, and SFB1064/A17 to H.L.]. J.R. is supported by a doctoral fellowship from Fonds de recherche du Québec - Santé. W.Q. was supported by the China Scholarship Council (CSC). We are grateful to the following colleagues for providing ESCs and somatic cell lines: Masahiro Muto and Haruhiko Koseki for mouse wt E14 and *Uhrf1*^{-/-} ESCs; En Li for mouse J1 wt; We thank David L. Spector for providing BHK cells containing a lac operator repeat array. M.S. is a fellow of the Integrated Research Training Group (IRTG) of the SFB1064. E.S. is a fellow of the International Max Planck Research School for Molecular and Cellular Life Sciences (IMPRS-LS) (GRK). We thank Georgia Kalideris for help with generation of the PAF15 KO cell line. We thank Linus Rinke and Andreas Meiser for assistance with super resolution microscopy experiments.

Conflict of interest

The authors declare no conflict of interest.

Author contributions

M.S., E.S. and H.L. conceived the study and wrote the manuscript. M.S. generated the *Uhrf* KO cell lines and performed confocal microscopy experiments. E.S. performed proteomics and biochemistry experiments and data analysis. I.F. and A.I. performed mass spectrometry measurements and contributed to experimental design and data analysis. W.Q. cloned the PAF15 constructs for ectopic expression. C.M. generated the PAF15 MIN tagged and KO cell line. J.R. performed and interpreted super resolution microscopy experiments.

References

- Achour, M., X. Jacq, P. Rondé, M. Alhosin, C. Charlot, T. Chataigneau, M. Jeanblanc, et al. 2008. "The Interaction of the SRA Domain of ICBP90 with a Novel Domain of DNMT1 Is Involved in the Regulation of VEGF Gene Expression." *Oncogene* 27 (15): 2187–97.
- Bostick, Magnolia, Jong Kyong Kim, Pierre-Olivier Estève, Amander Clark, Sriharsa Pradhan, and Steven E. Jacobsen. 2007. "UHRF1 Plays a Role in Maintaining DNA Methylation in Mammalian Cells." *Science* 317 (5845): 1760–64.
- Brown, Nicholas G., Ryan VanderLinden, Edmond R. Watson, Renping Qiao, Christy R. R. Grace, Masaya Yamaguchi, Florian Weissmann, et al. 2015. "RING E3 Mechanism for Ubiquitin Ligation to a Disordered Substrate Visualized for Human Anaphase-Promoting Complex." *Proceedings of the National Academy of Sciences of the United States of America* 112 (17): 5272–79.
- Citterio, Elisabetta, Roberto Papait, Francesco Nicassio, Manuela Vecchi, Paola Gomiero, Roberto Mantovani, Pier Paolo Di Fiore, and Ian Marc Bonapace. 2004. "Np95 Is a Histone-Binding Protein Endowed with Ubiquitin Ligase Activity." *Molecular and Cellular Biology* 24 (6): 2526–35.
- Cox, Jürgen, Marco Y. Hein, Christian A. Luber, Igor Paron, Nagarjuna Nagaraj, and Matthias Mann. 2014. "Accurate Proteome-Wide Label-Free Quantification by Delayed Normalization and Maximal Peptide Ratio Extraction, Termed MaxLFQ." *Molecular & Cellular Proteomics: MCP* 13 (9): 2513–26.
- Cox, Jürgen, and Matthias Mann. 2008. "MaxQuant Enables High Peptide Identification Rates, Individualized P.p.b.-Range Mass Accuracies and Proteome-Wide Protein Quantification." *Nature Biotechnology* 26 (12): 1367–72.
- Ding, Guangjin, Peilin Chen, Hui Zhang, Xiaojie Huang, Yi Zang, Jiwen Li, Jia Li, and Jiemin Wong. 2016. "Regulation of Ubiquitin-like with Plant Homeodomain and RING Finger Domain 1 (UHRF1) Protein Stability by Heat Shock Protein 90 Chaperone Machinery." *The Journal of Biological Chemistry* 291 (38): 20125–35.
- Du, Zhanwen, Jing Song, Yong Wang, Yiqing Zhao, Kishore Guda, Shuming Yang, Hung-Ying Kao, et al. 2010. "DNMT1 Stability Is Regulated by Proteins Coordinating Deubiquitination and Acetylation-Driven Ubiquitination." *Science Signaling* 3 (146): ra80.
- Easwaran, Hariharan P., Lothar Schermelleh, Heinrich Leonhardt, and M. Cristina Cardoso. 2004. "Replication-Independent Chromatin Loading of Dnmt1 during G2 and M Phases." *EMBO Reports* 5 (12): 1181–86.
- Harrison, Joseph S., Evan M. Cornett, Dennis Goldfarb, Paul A. DaRosa, Zimeng M. Li, Feng Yan, Bradley M. Dickson, et al. 2016. "Hemi-Methylated DNA Regulates DNA Methylation Inheritance through Allosteric Activation of H3 Ubiquitylation by UHRF1." *eLife* 5 (September). doi:10.7554/eLife.17101.
- Herce, Henry D., Wen Deng, Jonas Helma, Heinrich Leonhardt, and M. Cristina Cardoso. 2013. "Visualization and Targeted Disruption of Protein Interactions in Living Cells." *Nature Communications* 4: 2660.
- Hicke, Linda, and Rebecca Dunn. 2003. "Regulation of Membrane Protein Transport by Ubiquitin and Ubiquitin-Binding Proteins." *Annual Review of Cell and Developmental Biology* 19 (1): 141–72.
- Ho, S. N., H. D. Hunt, R. M. Horton, J. K. Pullen, and L. R. Pease. 1989. "Site-Directed Mutagenesis by Overlap Extension Using the Polymerase Chain Reaction." *Gene* 77 (1): 51–59.
- Huber, Wolfgang, Anja von Heydebreck, Holger Sültmann, Annemarie Poustka, and Martin Vingron. 2002. "Variance Stabilization Applied to Microarray Data Calibration and to the Quantification of Differential Expression." *Bioinformatics* 18 Suppl 1: S96–104.
- Jenkins, Yonchu, Vadim Markovtsov, Wayne Lang, Poonam Sharma, Denise Pearsall, Justin Warner, Christian Franci, et al. 2005. "Critical Role of the Ubiquitin Ligase Activity of UHRF1, a Nuclear RING Finger Protein, in Tumor Cell Growth." *Molecular Biology of the Cell* 16 (12): 5621–29.
- Liang, Chih-Chao, Bao Zhan, Yasunaga Yoshikawa, Wilhelm Haas, Steven P. Gygi, and Martin A. Cohn. 2015. "UHRF1 Is a Sensor for DNA Interstrand Crosslinks and Recruits FANCD2 to Initiate the Fanconi Anemia Pathway." *Cell Reports* 10 (12): 1947–56.
- Liu, Xiaoli, Qinqin Gao, Pishun Li, Qian Zhao, Jiqin Zhang, Jiwen Li, Haruhiko Koseki, and Jiemin Wong. 2013. "UHRF1 Targets DNMT1 for DNA Methylation through Cooperative Binding of Hemi-Methylated DNA and Methylated H3K9." *Nature Communications* 4: 1563.
- Luo, Tao, Shijun Cui, Chunjing Bian, and Xiaochun Yu. 2013. "Uhrf2 Is Important for DNA Damage Response in Vascular Smooth Muscle Cells." *Biochemical and Biophysical Research Communications* 441 (1): 65–70.
- Meilinger, Daniela, Karin Fellinger, Sebastian Bultmann, Ulrich Rothbauer, Ian Marc Bonapace, Wolfgang E. F. Klinkert, Fabio Spada, and Heinrich Leonhardt. 2009. "Np95 Interacts with de

- Novo DNA Methyltransferases, Dnmt3a and Dnmt3b, and Mediates Epigenetic Silencing of the Viral CMV Promoter in Embryonic Stem Cells." *EMBO Reports* 10 (11). EMBO Press: 1259–64.
- Mulholland, Christopher B., Martha Smets, Elisabeth Schmidtmann, Susanne Leidescher, Yolanda Markaki, Mario Hofweber, Weihua Qin, et al. 2015. "A Modular Open Platform for Systematic Functional Studies under Physiological Conditions." *Nucleic Acids Research* 43 (17): e112.
- Muto, M. 2002. "Targeted Disruption of Np95 Gene Renders Murine Embryonic Stem Cells Hypersensitive to DNA Damaging Agents and DNA Replication Blocks." *The Journal of Biological Chemistry* 277 (37): 34549–55.
- Nishiyama, Atsuya, Luna Yamaguchi, Jafar Sharif, Yoshikazu Johmura, Takeshi Kawamura, Keiko Nakanishi, Shintaro Shimamura, et al. 2013. "Uhrf1-Dependent H3K23 Ubiquitylation Couples Maintenance DNA Methylation and Replication." *Nature* 502 (7470): 249–53.
- Papait, Roberto, Christian Pistore, Diego Negri, Daniela Pecoraro, Lisa Cantarini, and Ian Marc Bonapace. 2007. "Np95 Is Implicated in Pericentromeric Heterochromatin Replication and in Major Satellite Silencing." *Molecular Biology of the Cell* 18 (3): 1098–1106.
- Pickart, Cecile M. 2004. "Back to the Future with Ubiquitin." *Cell* 116 (2): 181–90.
- Pickart, Cecile M., and David Fushman. 2004. "Polyubiquitin Chains: Polymeric Protein Signals." *Current Opinion in Chemical Biology* 8 (6): 610–16.
- Povlsen, Lou K., Petra Beli, Sebastian A. Wagner, Sara L. Poulsen, Kathrine B. Sylvestersen, Jon W. Poulsen, Michael L. Nielsen, Simon Bekker-Jensen, Niels Mailand, and Chunaram Choudhary. 2012. "Systems-Wide Analysis of Ubiquitylation Dynamics Reveals a Key Role for PAF15 Ubiquitylation in DNA-Damage Bypass." *Nature Cell Biology* 14 (10): 1089–98.
- Qin, Weihua, Heinrich Leonhardt, and Fabio Spada. 2011. "Usp7 and Uhrf1 Control Ubiquitination and Stability of the Maintenance DNA Methyltransferase Dnmt1." *Journal of Cellular Biochemistry* 112 (2): 439–44.
- Qin, Weihua, Patricia Wolf, Nan Liu, Stephanie Link, Martha Smets, Federica La Mastra, Ignasi Forné, et al. 2015. "DNA Methylation Requires a DNMT1 Ubiquitin Interacting Motif (UIM) and Histone Ubiquitination." *Cell Research* 25 (8): 911–29.
- Quenneville, Simon, Gaetano Verde, Andrea Corsinotti, Adamandia Kapopoulou, Johan Jakobsson, Sandra Offner, Ilaria Baglivo, et al. 2011. "In Embryonic Stem Cells, ZFP57/KAP1 Recognize a Methylated Hexanucleotide to Affect Chromatin and DNA Methylation of Imprinting Control Regions." *Molecular Cell* 44 (3): 361–72.
- R Core Team. 2016. "R: A Language and Environment for Statistical Computing." Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Ritchie, Matthew E., Belinda Phipson, Di Wu, Yifang Hu, Charity W. Law, Wei Shi, and Gordon K. Smyth. 2015. "Limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies." *Nucleic Acids Research* 43 (7): e47.
- Rottach, Andrea, Elisabeth Kremmer, Danny Nowak, Prisca Boisguerin, Rudolf Volkmer, M. Cristina Cardoso, Heinrich Leonhardt, and Ulrich Rothbauer. 2008. "Generation and Characterization of a Rat Monoclonal Antibody Specific for PCNA." *Hybridoma* 27 (2): 91–98.
- Sharif, Jafar, Masahiro Muto, Shin-Ichiro Takebayashi, Isao Suetake, Akihiro Iwamatsu, Takaho A. Endo, Jun Shinga, et al. 2007. "The SRA Protein Np95 Mediates Epigenetic Inheritance by Recruiting Dnmt1 to Methylated DNA." *Nature* 450 (7171): 908–12.
- Smyth, Gordon K. 2004. "Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments." *Statistical Applications in Genetics and Molecular Biology* 3 (February): Article3.
- Solovei, Irina, Solovei Irina, and Cremer Marion. 2010. "3D-FISH on Cultured Cells Combined with Immunostaining." In *Methods in Molecular Biology*, 117–26.
- Sun, Lijun, and Zhijian J. Chen. 2004. "The Novel Functions of Ubiquitination in Signaling [Curr. Opin. Cell Biol. 16 (2004) 119]." *Current Opinion in Cell Biology* 16 (3): 339–40.
- Tian, Yanyan, Manikandan Paramasivam, Gargi Ghosal, Ding Chen, Xi Shen, Yaling Huang, Shamima Akhter, et al. 2015. "UHRF1 Contributes to DNA Damage Repair as a Lesion Recognition Factor and Nuclease Scaffold." *Cell Reports* 10 (12): 1957–66.
- Tsukamoto, T., N. Hashiguchi, S. M. Janicki, T. Tumber, A. S. Belmont, and D. L. Spector. 2000. "Visualization of Gene Activity in Living Cells." *Nature Cell Biology* 2 (12): 871–78.
- Tyanova, Stefka, Tikira Temu, Pavel Sinitcyn, Arthur Carlson, Marco Y. Hein, Tamar Geiger, Matthias Mann, and Jürgen Cox. 2016. "The Perseus Computational Platform for Comprehensive Analysis of (prote) Omics Data." *Nature Methods* 13 (9). Nature Research: 731–40.
- Udeshi, Namrata D., Philipp Mertins, Tanya Svinkina, and Steven A. Carr. 2013. "Large-Scale Identification of Ubiquitination Sites by Mass Spectrometry." *Nature Protocols* 8 (10): 1950–60.
- Uemura, T., E. Kubo, Y. Kanari, T. Ikemura, K. Tatsumi, and M. Muto. 2000. "Temporal and Spatial

- Localization of Novel Nuclear Protein NP95 in Mitotic and Meiotic Cells." *Cell Structure and Function* 25 (3): 149–59.
- UniProt Consortium. 2015. "UniProt: A Hub for Protein Information." *Nucleic Acids Research* 43 (Database issue): D204–12.
- Xie, Chanlu, Mu Yao, and Qihan Dong. 2014. "Proliferating Cell Nuclear Antigen-Associated Factor (PAF15): A Novel Oncogene." *The International Journal of Biochemistry & Cell Biology* 50 (May): 127–31.
- Xu, Guoqiang, Jeremy S. Paige, and Samie R. Jaffrey. 2010. "Global Analysis of Lysine Ubiquitination by Ubiquitin Remnant Immunoaffinity Profiling." *Nature Biotechnology* 28 (8): 868–73.
- Yaffe, D., and O. Saxel. 1977. "Serial Passaging and Differentiation of Myogenic Cells Isolated from Dystrophic Mouse Muscle." *Nature* 270 (5639): 725–27.
- Yu, Peiwen, Betty Huang, Mary Shen, Clorinda Lau, Eva Chan, Jennifer Michel, Yue Xiong, Donald G. Payan, and Ying Luo. 2001. "p15PAF, a Novel PCNA Associated Factor with Increased Expression in Tumor Tissues." *Oncogene* 20 (4): 484–89.
- Zhang, Haoxing, Hailong Liu, Yali Chen, Xu Yang, Panfei Wang, Tongzheng Liu, Min Deng, et al. 2016. "A Cell Cycle-Dependent BRCA1-UHRF1 Cascade Regulates DNA Double-Strand Break Repair Pathway Choice." *Nature Communications* 7 (January): 10201.

Figures and Figure legends

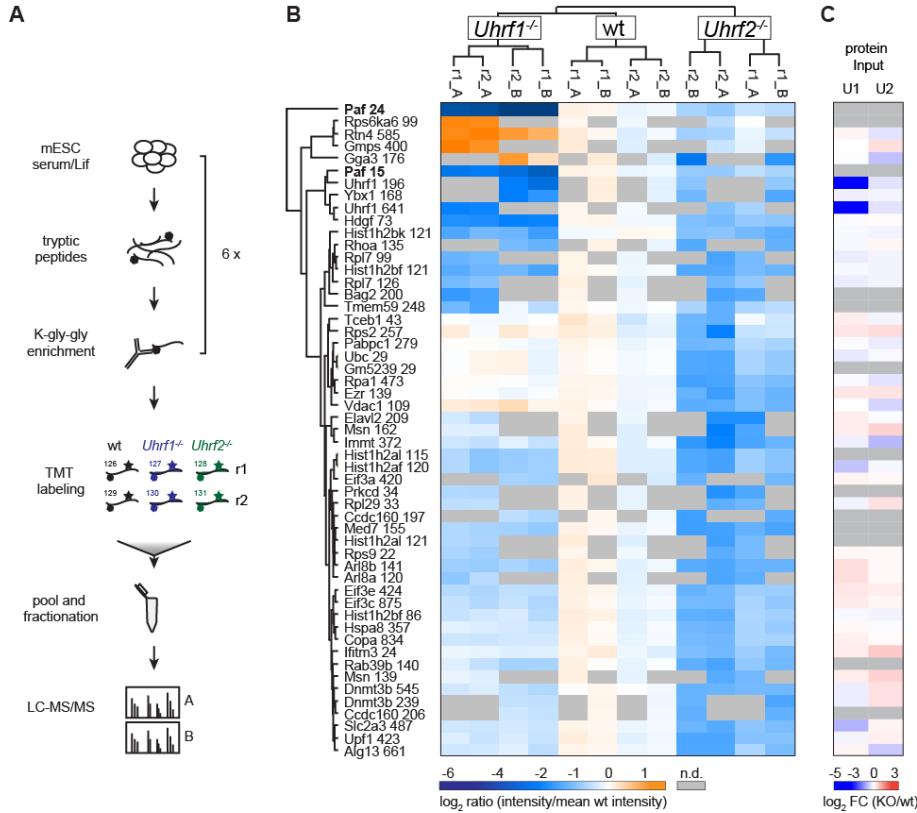


FIGURE 1. Ubiquitome characterization of mouse embryonic stem cells (ESCs) deficient for UHRF1 and UHRF2. (A) Experimental workflow. *Uhrf1*^{-/-}, *Uhrf2*^{-/-} and wt mouse ESCs were digested to peptides and ubiquitin remnant motif (K-gly-gly) -containing peptides were enriched using an antibody. Peptides were labelled using TMT sixplex reagents, pooled for fractionation and subsequent mass spectrometry analysis. (B) Heatmap of differentially ubiquitinated K-gly-gly peptides (Gene name _ amino acid position of ubiquitination) identified in wt, *Uhrf1*^{-/-} and *Uhrf2*^{-/-} ESCs. Only peptides with at least a three-fold intensity change ($\log_2 > 1.58$) in at least two replicates are shown (53 peptides out of total 1248). Experiments were carried out in biological (r1, r2) and technical duplicates (_A, _B). (C) Total protein abundance (\log_2 LFQ intensity fold change) of the respective peptides in *Uhrf1*^{-/-} (U1) and *Uhrf2*^{-/-} (U2) relative to wt cells.

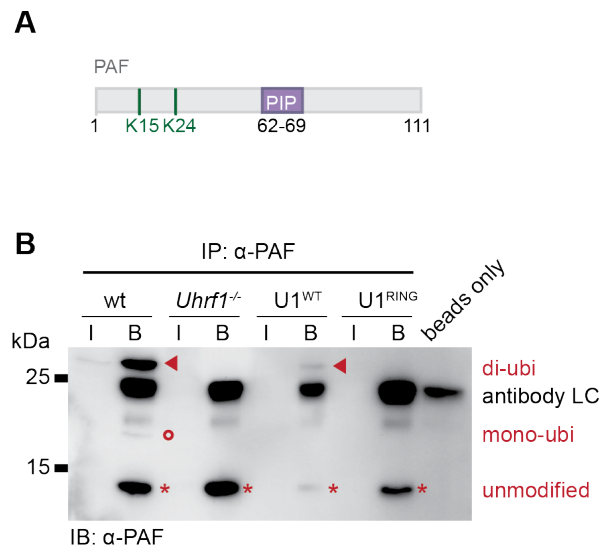


FIGURE 3. PAF15 ubiquitination by UHRF1. (A) Schematic outline of the PAF15 protein. (B) Co-immunoprecipitation and Western Blot analysis of endogenous PAF15 from wt (E14), $Uhrf1^{-/-}$ and $Uhrf1^{-/-}$ ESC expressing wt UHRF1-GFP ($U1^{WT}$) and RING domain mutant (H730A) construct ($U1^{RING}$). Antibody conjugated beads were used as negative control. I = Input, B = Bound. Asterisks indicate unmodified PAF15 while circles and triangles indicate mono- and di-ubiquitinated PAF15, respectively.

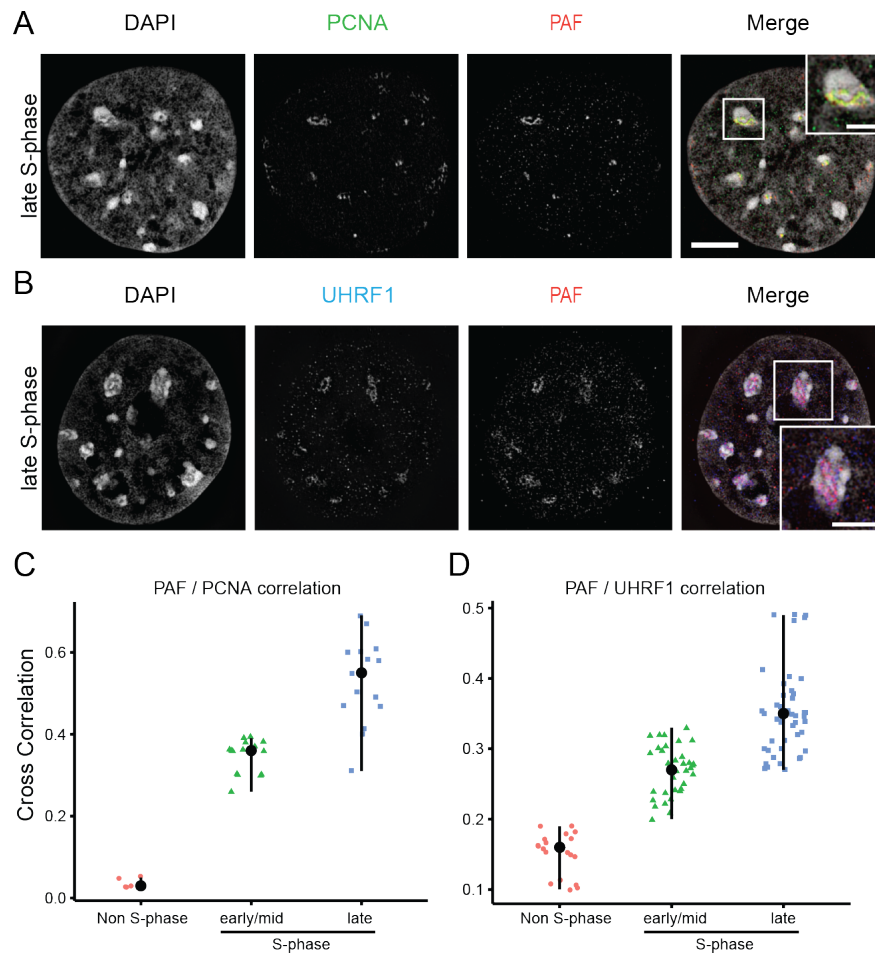


FIGURE 4. PAF15 localization with PCNA and UHRF1 throughout cell cycle. (A+B) 3D-SIM nuclear mid-sections of anti-PAF15 (red) antibody distributions with (A) anti-PCNA (green) and (B) anti-UHRF1 (green) with DAPI counterstaining (gray) in C2C12 cells. Scale bar = 5 μm and 2 \times magnifications of selected boxed regions. Scale bars = 2,5 μm . (C) Pearson correlation coefficient of PAF15 and PCNA (C) and PAF15 and UHRF1 (D) in non S-phase, early/mid and late S-phase C2C12 cells depicted as scatter plots with median and 95% confidence interval.

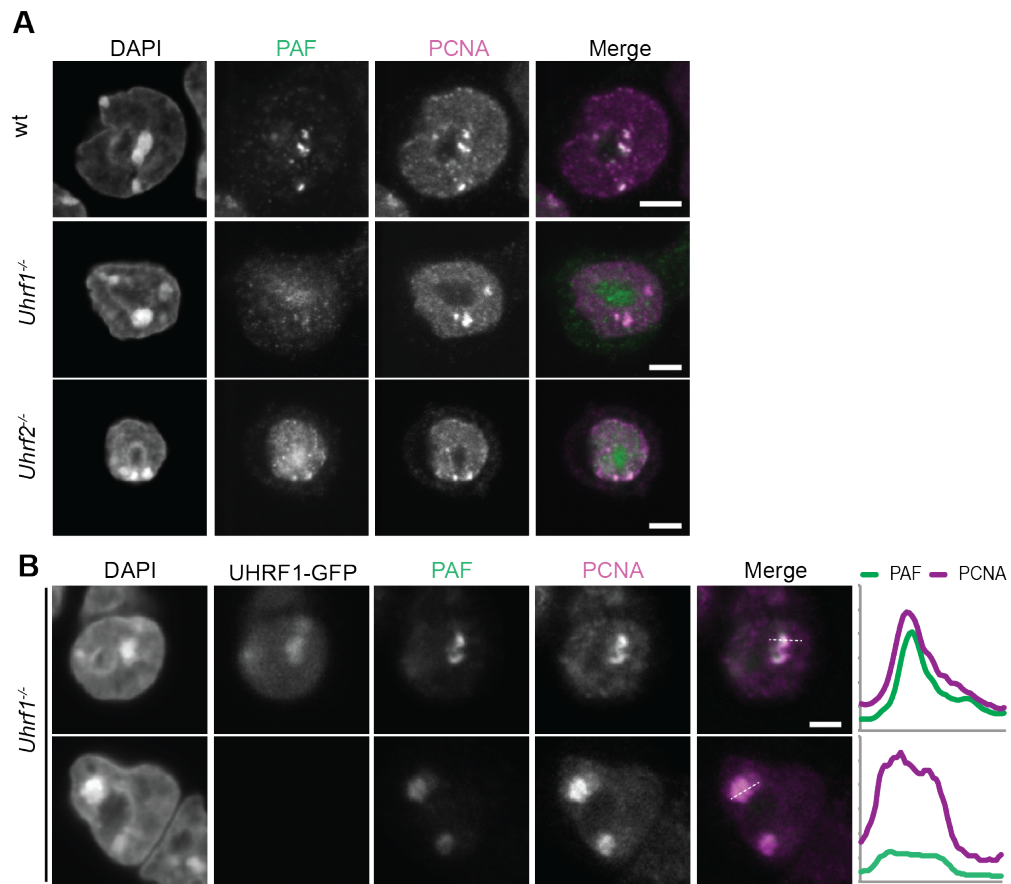


Figure 5. PAF15 localization in dependence of UHRF1. (A) Confocal mid sections of wt, *Uhrf1*^{-/-} and *Uhrf2*^{-/-} ESC were stained with antibodies anti-PAF15 (green) and PCNA (magenta). DNA was counterstained with DAPI. (B) Confocal mid sections of *Uhrf1*^{-/-} ESCs expressing UHRF1-GFP. ESCs were stained with antibodies anti-PAF15 (green) and PCNA (magenta). DNA was counterstained with DAPI. Scale bars = 5 μ m. Line intensity profiles of PAF15 and PCNA are shown next to the image.

SUPPLEMENTARY DATA

Table S1 (corresponding to Figure 1)

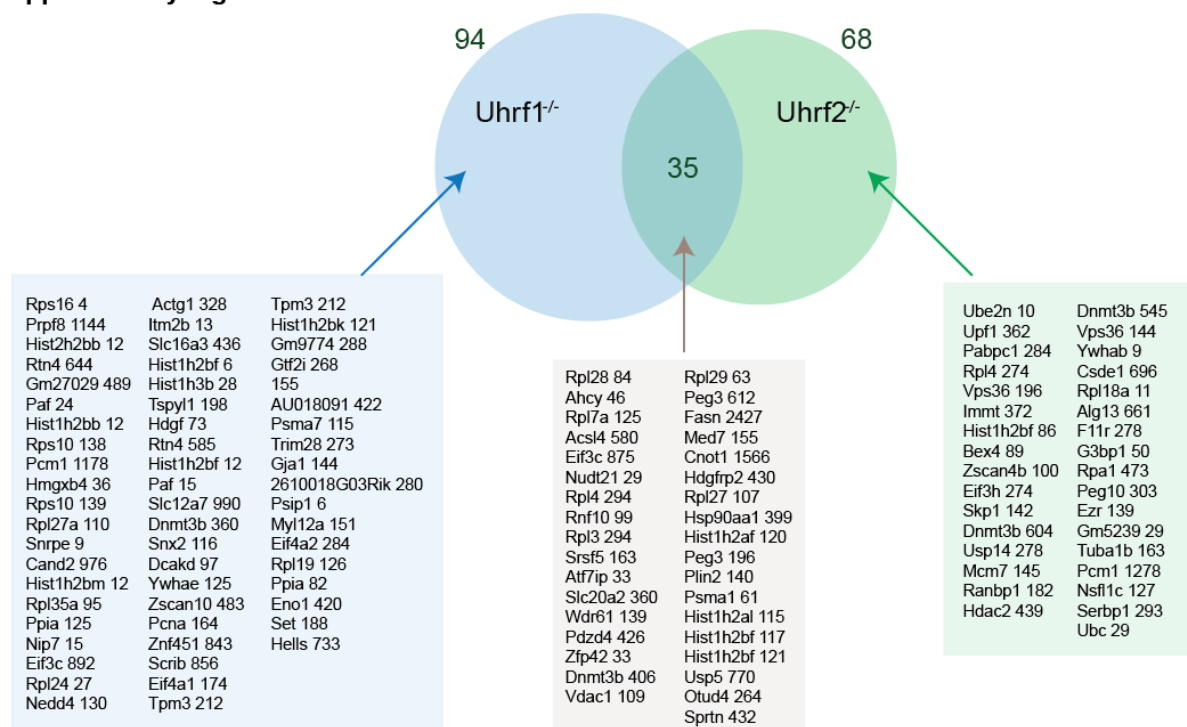
Quantified K-gly-gly peptides in wildtype (wt), Uhrf1^{-/-} (U1) and Uhrf2^{-/-} (U2) embryonic stem cells.

Table S2 (corresponding to Figure 2)

Limma analysis results of differentially regulated K-gly-gly peptides in Uhrf1^{-/-} and Uhrf2^{-/-} embryonic stem cells.

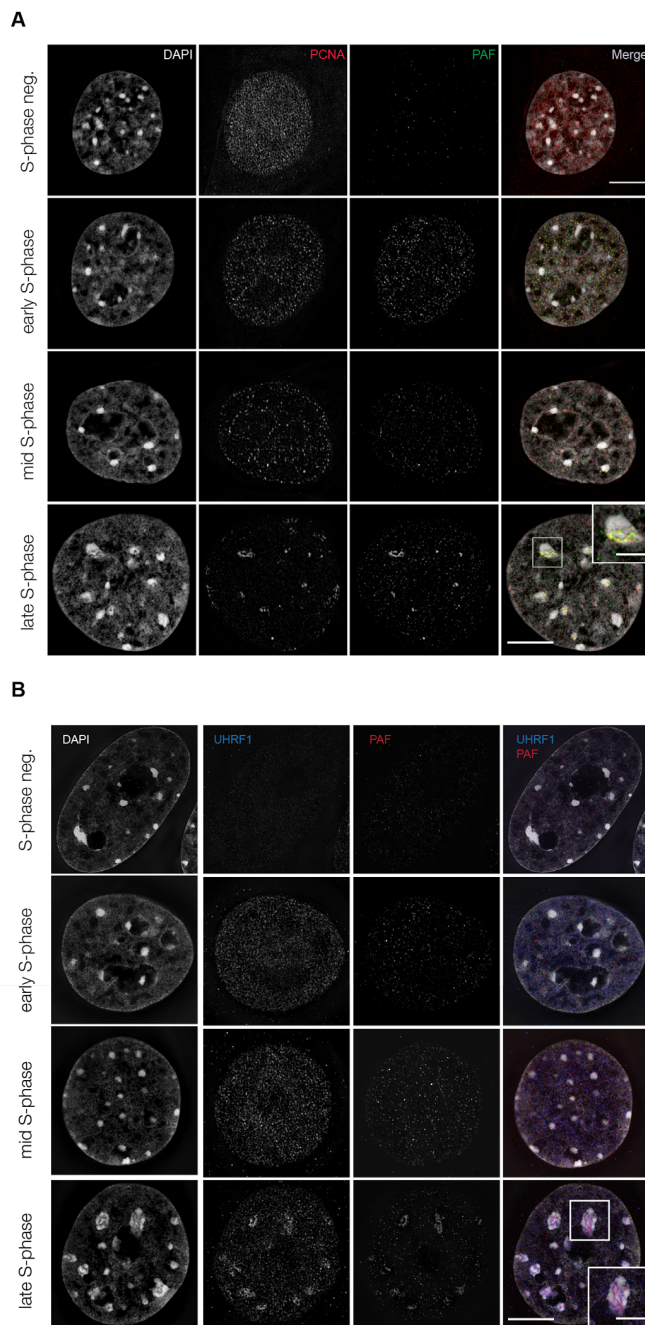
Figure S1

Supplementary Figure S1



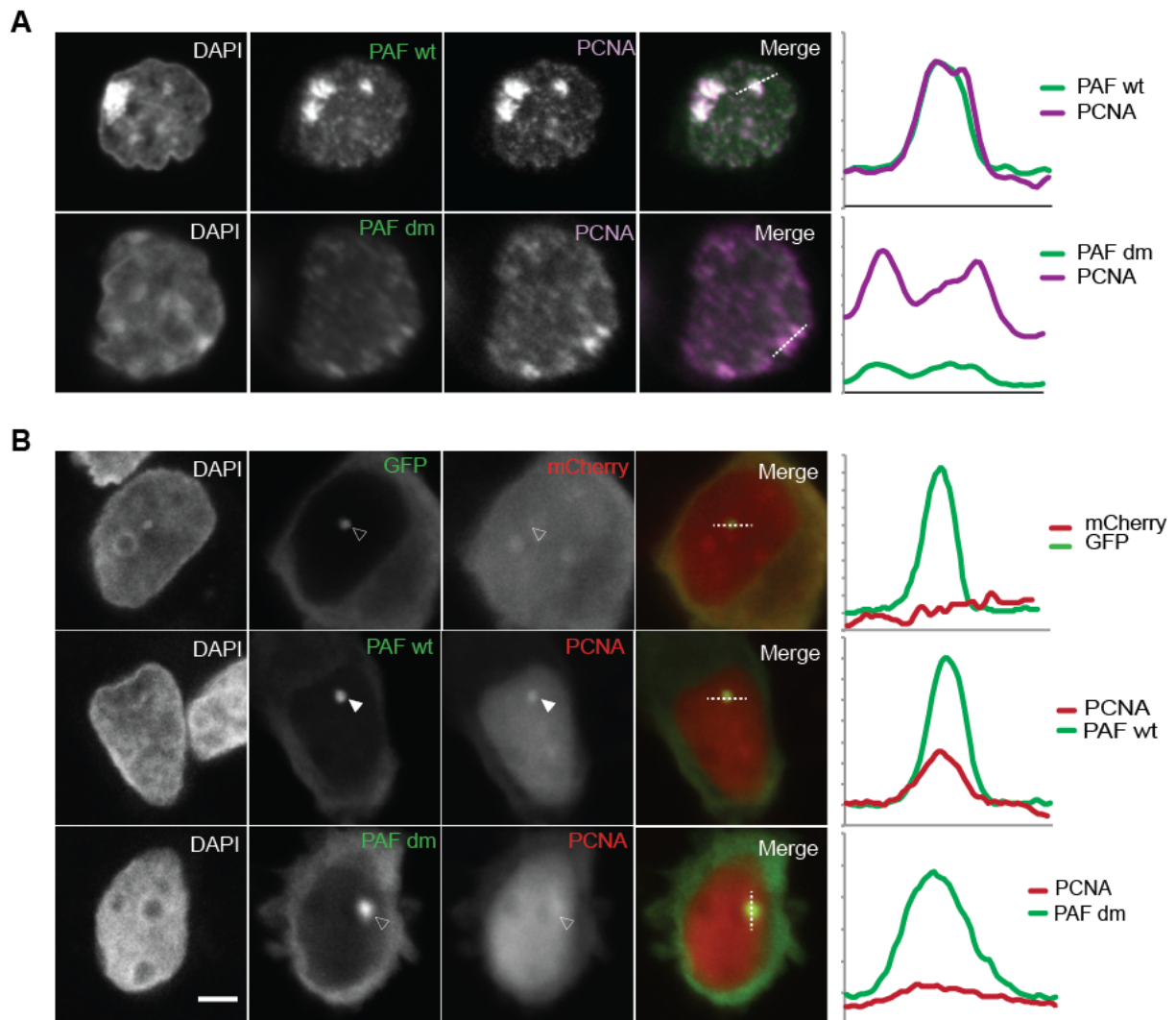
Supplementary Figure S1: Venn diagram of significantly regulated K-gly-gly peptides (gene name_position of ubiquitination) in Uhrf1^{-/-} and Uhrf2^{-/-} cells (Limma adj. p-value < 0.05).

Figure S2



Supplementary Figure S2: PAF co-localization with PCNA and UHRF1 is cell cycle dependent. (A) 3D-SIM nuclear mid-sections of anti-PCNA (green) and anti-PAF (red) antibody distributions with DAPI counterstaining (gray) in C2C12 cells in different stages of S-phase. (B) 3D-SIM nuclear mid-sections of anti-UHRF1 (blue) and anti-PAF (red) antibody distributions with DAPI counterstaining (gray) in C2C12 cells. Scale bar = 5 μ m and 2 \times magnifications of selected boxed regions. Scale bar = 2,5 μ m

Figure S3



Supplementary Figure S3: Di-ubiquitinated PAF is recruited to PCNA. (A) Paf KO mESCs rescued with transiently transfected with GFP-PAF wild type (wt, upper panel) or GFP-PAF K15R/K24R mutant (dm, lower panel). Line intensity profiles of PAF and PCNA are shown next to the image. (B) Analysis of ubiquitination-mediated recruitment of PAF to PCNA in a cell-based F3H assay. F3H assay of mCherry-PCNA (red) with GFP-PAF wildtype and double mutant GFP-PAF K15R/K24R (green). Line intensity profiles of the GFP-PAF constructs and mCherry-PCNA are shown next to the images. Scale bar = 5 μ m.

3 Discussion

3.1 DNMT1 mutations found in HSAN1E patients affect interaction with UHRF1 and neuronal differentiation

DNA methylation has been regarded a stable epigenetic mark set by the *de novo* DNA methyltransferases DNMT3A and DNMT3B during development and maintained by DNMT1 after DNA replication (Bird 2002; Goll and Bestor 2005; X. Cheng and Blumenthal 2008; Law and Jacobsen 2010; Smith and Meissner 2013). However, this view of DNMT1 as a simple copy machine, cannot explain why DNMT1 is expressed and required in postmitotic cells like neurons (Goto et al. 1994; Inano et al. 2000). Despite or because of the central and ubiquitous role in DNA methylation, only very few human diseases were linked to *DNMT1* mutations. To date, 14 different *DNMT1* point mutations and one deletion mutation have been identified, which are associated with two neurodegenerative diseases, HSAN1E and ADCA-DN (4.3 Supplementary Material, Supplementary Table) (Klein et al. 2011; Rotthier et al. 2012; Winkelmann et al. 2012). Strikingly, all point and deletion mutations are located within the TS domain in the NTD of DNMT1 (Figure 10), which is crucially involved in the functional regulation of DNMT1 activity (Leonhardt et al. 1992). HSAN1E associated mutations cluster in the central part, while ADCA-DN associated mutations are located in the C-terminal part of the TS domain. However, the underlying molecular mechanism of DNMT1 dysfunction remains under investigation.

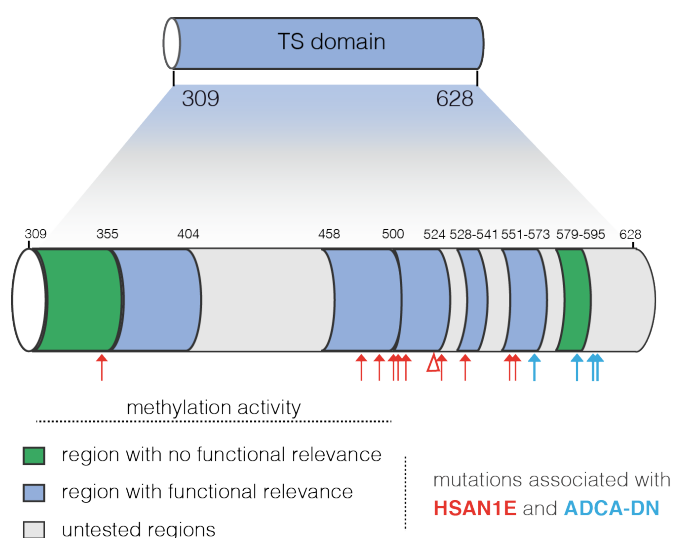


Figure 10. The TS domain of DNMT1. Outline of the TS domain of DNMT1 and the enzymatically relevant regions within the TS domain as determined by functional complementation assays (Smets et al. 2017). HSAN1E (red) and ADCA-DN (blue) associated point mutations (Klein et al. 2011; Rotthier et al. 2012; Winkelmann et al. 2012) are indicated by arrows and the deletion by Δ.

To investigate how a single exchange in the amino acid sequence outside the catalytic domain of DNMT1 may lead to global hypomethylation and neurodegenerative diseases, we transferred two mutations (P496Y and Y500C) to an ESC differentiation system and assayed for cellular defects. As UHRF1 is an essential cofactor of DNMT1 and the TS domain in DNMT1 has been described to mediate the interaction with UHRF1 (Achour et al. 2008; Felle et al. 2011), we first investigated whether the HSAN1E mutations affect this protein-protein interaction. Our co-immunoprecipitation experiments and deletion analyses are consistent with previous yeast-two hybrid screens (Achour et al. 2008) showing an interaction of the TS domain with the SRA domain of UHRF1. In absence of hemimethylated DNA, the TS domain binds the catalytic center in the CTD of DNMT1 and inhibits *de novo* methylation. This autoinhibitory mechanism is resolved when the TS domain interacts with the SRA domain of UHRF1 resulting in allosteric activation of DNMT1 (Bashtrykov, Jankevicius, et al. 2014; Berkyurek et al. 2014). Consequently, the release of the TS domain from the catalytic center of DNMT1 accompanied by a conformational change is a prerequisite for enzyme activation (Takeshita et al. 2011). Interestingly, HSAN1E mutations are located within the part of the TS domain that interacts with UHRF1 and weaken the protein-protein interaction.

We could further demonstrate that mutated DNMT1 failed to maintain DNA methylation in ESCs and showed weaker association with heterochromatin in late S phase, which is consistent with the previous observation that UHRF1 is required for recruitment of DNMT1 (Bostick et al. 2007). The weakened interaction of DNMT1 with UHRF1 could explain the reduced activity of mutated DNMT1 pull-downs *in vitro* (Klein et al. 2011) and fits well with our observation that DNMT1 harboring HSAN1E mutations is impaired in catalytic complex formation as measured with our trapping assay in living cells (Schermelleh et al. 2005).

In addition to defects in enzyme activation and targeting, we show that HSAN1E mutations render DNMT1 unstable in late S and G2 phase due to proteasomal degradation. This might be a result of failed heterochromatin binding, which is in line with previous studies (Klein et al. 2011). Chromatin unbound DNMT3A and DNMT3B showed a similar protein destabilization (Jeong et al. 2009; Sharma et al. 2011). In addition to UHRF1 binding, DNMT1 activity is influenced by further protein interactions and multiple posttranslational modifications, which likely contribute to proper regulation of maintenance DNA methylation (Qin, Leonhardt, and Pichler 2011).

Several of these interactions may not be absolutely required but may enhance efficiency and thereby contribute to fine tuning of DNMT1 activity as we found for the interaction with PCNA, which two-fold enhances DNMT1 efficiency in living cells (Schermelleh et al. 2007; Spada et al. 2007). While it is unclear which of

Discussion

these interactions are impaired by the HSN1E mutations in different cell types, we could clearly show that ESCs carrying these mutations are impaired in neuronal differentiation, which is accompanied by an increased apoptosis rate.

Interestingly, although HSN1E mutations in the TS domain are present in all patient cells, they mostly affect the neuronal lineage and induce both central and peripheral neurodegeneration. Although it is not clear yet how dynamic DNA methylation changes could be in postmitotic neurons, the turnover of DNA methylation was proposed as one possible modulator of neuronal plasticity in response to external or internal stimuli (Fan et al. 2001; Borrelli et al. 2008; Yu, Baek, and Kaang 2011).

This fits well with the observation that neuronal tissues are characterized by high 5hmC levels and strong expression of TET genes (Tahiliani et al. 2009; Globisch et al. 2010; Szwagierczak et al. 2010; Ito et al. 2011). Therefore, DNA modifications might be more dynamic in postmitotic neurons than previously thought so that even subtle changes in protein-protein interactions and enzymatic activity might unbalance the equilibrium of DNA modifications. However, it remains unclear why in particular neuronal tissues are affected by these *DNMT1* mutations in HSN1E patients.

HSN1E patients are characterized by heterozygous mutations (Klein et al. 2011). However, Sun and colleagues only detected subtle changes in DNA methylation in HSN1E patients with whole-genome bisulfite sequencing (Z. Sun et al. 2014). Our experimental setup is based on a stable complementation experiment in a *DNMT1*^{-/-} background, which is suited to amplify the mutant phenotype for biochemical studies. To gain insight into the biological pathways underlying the late-onset degeneration of neuronal tissues and for physiological relevance, future studies might benefit from mouse models that can help to understand the molecular dysfunction. Moreover, a promising research model for functional assays to further understand HSN1E related mechanisms under physiological conditions might be the use of human induced pluripotent stem cells (iPSCs) derived from patients (G. Lee et al. 2009; Jung et al. 2012).

In summary, we show that HSN1E mutations affect the direct DNMT1 interaction with the essential maintenance DNA methylation cofactor UHRF1, cause cell cycle dependent degradation of DNMT1 and impair neuronal differentiation. These data add to our understanding of the role and regulation of DNMT1 during differentiation and help to understand DNMT1 dysfunction and hypomethylation in the pathogenesis of this neurodegenerative disease.

3.2 DNA methylation requires a DNMT1 ubiquitin interacting motif (UIM) and histone ubiquitination

The maintenance DNA methyltransferase DNMT1 specifically recognizes hemimethylated DNA substrates. However, the preference of DNMT1 for hemimethylated DNA measured *in vitro* (Bestor and Ingram 1983; Yoder et al. 1997; Jeltsch 2006; Frauer and Leonhardt 2009) is not sufficient to explain efficient maintenance of DNA methylation patterns over many cell division cycles *in vivo*. As already discussed in 3.1, the cofactor UHRF1 is essential for DNA methylation as UHRF1 recruits and allosterically activates DNMT1 (Sharif et al. 2007; Bostick et al. 2007; Bashtrykov, Jankevicius, et al. 2014; Berkyurek et al. 2014). For this process, heterochromatin binding of UHRF1 is mediated by the TTD, PHD and SRA domain and defects in any of these domains lead to lower DNA methylation levels (Jeltsch 2006; Rothbart et al. 2012, 2013). Accordingly, it was suggested that UHRF1 reads and binds hemimethylated DNA and repressive histone marks and then by binding DNMT1 recruits it for maintenance DNA methylation. Interestingly, the RING domain of UHRF1 was shown to ubiquitinate histones but the function of his catalytic activity was unknown (Citterio et al. 2004).

In this study, I participated in investigating the distinct roles of different UHRF1 domains and their function in mediating DNA methylation. We found that the RING domain is important for DNA methylation by DNMT1, although it is not required for UHRF1 chromatin binding or interaction with DNMT1. However, UHRF1 with decreased ubiquitin E3 ligase activity failed to recruit DNMT1 to chromatin suggesting that DNMT1 recruitment to replication forks is influenced by the ubiquitination activity of the RING domain.

Recently, it was shown that ubiquitination of H3K23 in *Xenopus* egg extracts by UHRF1 provides a binding site for DNMT1 on chromatin. This mark recruits DNMT1 and further enables maintenance DNA methylation throughout replication (Nishiyama et al. 2013). Nishiyama et al. showed that a 100 bp TS domain deletion mutant does not bind ubiquitinated histone H3 *in vitro*, in contrast to the wt TS domain (Nishiyama et al. 2013). To understand how the TS domain may contribute to H3 ubiquitination binding, we screened for potential binding motifs within DNMT1 using bioinformatic analyses and found a ubiquitin interacting motif (UIM) in the N-terminal part of the TS domain of DNMT1.

Further, we identified H3K18 as the ubiquitination substrate of UHRF1 in mammalian cells using proteomic analysis and could show that the UIM mediates the recognition of ubiquitinated H3 *in vitro*. The binding of ubiquitinated H3K18 (H3K18ub) is essential for DNMT1 subnuclear distribution and maintenance DNA

Discussion

methylation. Moreover, we confirmed that hemimethylated DNA is bound by the SRA domain (Bostick et al. 2007; Sharif et al. 2007) and could show that UHRF1 PHD binding to H3R2 is a prerequisite for H3 ubiquitination. Taken together, we propose that concomitant binding of the TTD, the PHD and the SRA domain to H3K9me3, H3R2 and hemimethylated DNA, respectively, is required for H3K18 ubiquitination and further DNA methylation.

Recently, DNMT1 was identified as a specific binder of H2Aub nucleosomes in mammalian cells (Kalb et al. 2014). Consistently, we found that DNMT1 binds ubiquitinated H2AK119 (H2AK119ub) via its UIM. H2A ubiquitination is set by the dimer RING1A and RING1B as part of PRC1 and was shown to be important in regulating gene expression (Leeb and Wutz 2007). Similar to UHRF1-dependent H3 ubiquitination, H2AK119ub might play a role in DNA methylation. We propose that UIM-mediated binding of DNMT1 to H2AK119ub might direct DNMT1 to un- or hemimethylated sites dependent on PRC1 ubiquitination activity.

In absence of its enzymatic substrate and UHRF1, DNMT1 activity is conformationally inhibited via an intramolecular interaction between its TS domain and the catalytic domain (see 1.3.1) (Syeda et al. 2011; Takeshita et al. 2011). Based on our observations, we speculate that the UIM may also play a role in the activation of the enzyme. In such a scenario, the interaction between UIM with ubiquitinated histone tails outcompetes the autoinhibitory TS/CD interaction and induces a conformational change, allosterically activating the enzyme

Ubiquitination can be reversed by USPs. As the UHRF1-DNMT1 complex interacts with USP7, which was shown to deubiquitinate and thereby stabilize DNMT1 (Du et al. 2010; Qin, Leonhardt, and Spada 2011), USP7 might regulate histone H3 ubiquitination and thereby DNMT1 targeting. Interestingly, among other core histone lysine residues, H3K18 was found to be acetylated (Kurdistani, Tavazoie, and Grunstein 2004; Tsai et al. 2010), giving rise to an alternative pathway to regulate H3K18ub abundance and thereby DNMT1 chromatin association. Acetylated H3K18 is enriched at TSSs of active and poised genes (Zhibin Wang et al. 2008) and might prevent the binding and silencing by DNMT1. In consequence, proteins that might be involved in the transcriptional activation and repression via the regulation of H3K18ac abundance are of major interest. Promising candidates might be the sirtuin (SIRT) family members as they were shown to deacetylate histones as well as non-histone proteins (Chalkiadaki and Guarente 2015). SIRTs are conserved enzymes that catalyze NAD⁺-dependent deacetylation reactions with a broad range of substrates such as tumor suppressor p53, forkhead transcription factors, nuclear receptor co-activator PGC-1 α , histone acetyltransferase p300, and NF κ B (Mantel and Broxmeyer 2008). In mammals, there are seven homologue SIRTs: SIRT1, SIRT2, SIRT6 and SIRT7 localize in the

nucleus, while SIRT3, SIRT4 and SIRT5 are found in the mitochondria. SIRT6 has been shown to bind to bivalent genomic regions, to affect DNA methylation by modulating DNMT1 as well as to modify the histone acetylation status of the *DNMT3L* promoter (Kuzmichev et al. 2005; Peng et al. 2011; Heo et al. 2017). These results suggest promising candidate proteins to investigate histone deacetylation and subsequent DNA methylation dynamics (Figure 11).

Finally, H3K18 has also been found to be methylated (Zee et al. 2010) by yet unknown HMTs suggesting another possible regulatory pathway that might influence DNMT1 recruitment or repulsion.

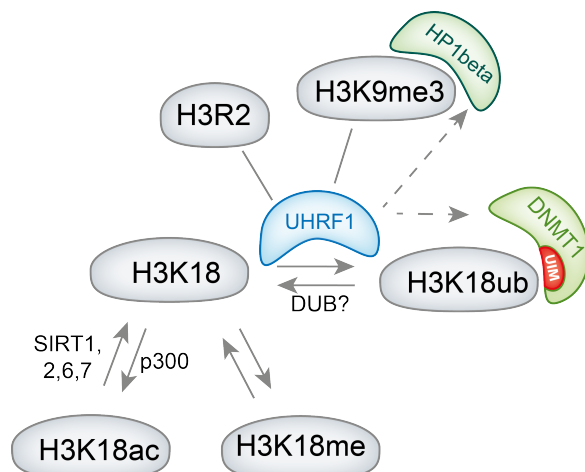


Figure 11. Overview of dynamic H3 tail modifications with known and postulated protein interactions involved in the regulation of DNA modifications. Recognition of H3K18ub via the UIM of DNMT1 is essential for DNA methylation. Therefore, the (de-) ubiquitination of H3K18 and competing PTMs like acetylation and methylation are candidates for indirect regulation of DNA methylation.

In summary, we showed that DNA methylation by DNMT1 requires recognizing of H3R2-K9me3 histone tails and hemimethylated DNA by UHRF1, which then ubiquitinates H3K18 with its RING domain and thereby constitutes a binding site for the UIM of DNMT1. Dynamic histone modifications likely control DNMT1 chromatin binding and thereby direct methylation activity. As ubiquitination of H3K18 is required for DNA methylation, it is now of interest how this ubiquitination mark is removed and how competing PTMs of H3K18 like (de)acetylation and (de)methylation regulate the availability of H3K18 for ubiquitination with subsequent DNA methylation.

3.3 A modular open platform for systematic functional studies under physiological conditions

CRISPR/Cas has revolutionized the genome engineering field and provides an easy and fast method for the site-directed manipulation of genes and genomes in cells and organisms. While genome-wide gene disruption screenings are informative about genes and pathways involved (Shalem et al. 2014; T. Wang et al. 2014), further investigation and functional characterization of target genes is essential to understand their function. It is important that biochemical, genetic and cell biological studies are applied under physiological conditions that require complex and time-consuming genetic manipulations.

CRISPR/Cas mediated DSBs can either lead to indel (insertions or deletions) formation due to error-prone NHEJ or to HDR. Together with short homology donors, HDR allows targeted sequence insertion into the genome (Capecchi 2005). However, a significant risk of off-target cleavage mediated by CRISPR/Cas can also lead to NHEJ and thereby result in indel formation (Kuscu et al. 2014; X. Wu et al. 2014). Due to this off-target activity, CRISPR/Cas in a classical sense can lead to unwanted gene disruption and might confound the use in research and therapeutic applications. In this project, I participated in establishing a more efficient approach (multifunctional integrase (MIN)-tag strategy) (Figure 12) that offers rapid, precise and efficient genetic manipulation of target genes.

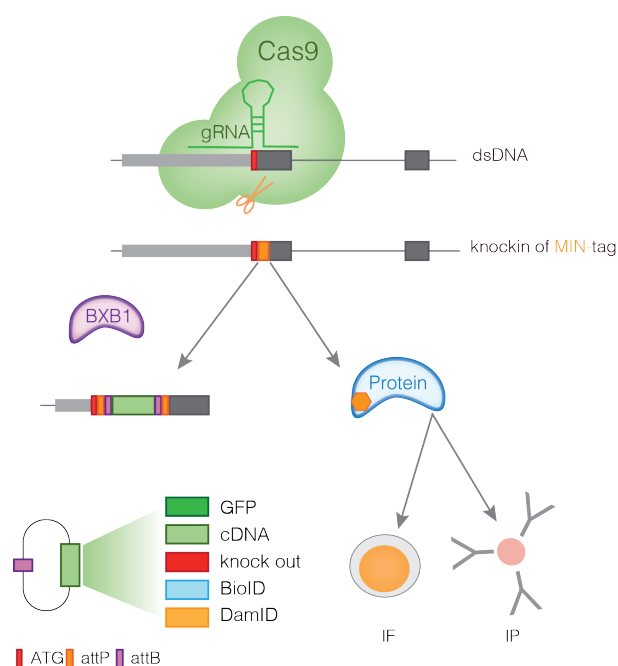


Figure 12. Schematic outline of the MIN tag strategy. A serine integrase site (attP) is inserted in-frame after the start codon mediated by CRISPR/Cas assisted HR resulting in a novel epitope tag that can be used for downstream experiments *in vitro*. The attP site can then be used to integrate functional cassettes on the genetic level. Modified from (Mulholland et al. 2015).

Discussion

The MIN tag strategy reduces the risk of off-target effects as it only requires a single CRISPR/Cas mediated cleavage event to insert a donor sequence (MIN tag) that carries a recognition site (attP) for BXB1, a highly specific integrase (Keravala et al. 2006; Russell et al. 2006; Brown et al. 2011; Z. Xu et al. 2013). The genomic sequence of the MIN tag enables BXB1-mediated recombination of any prefabricated functional cassettes that are then expressed under the respective, endogenous promoter (Figure 12).

Fusing the MIN tag in frame with the desired target gene results in the expression of a novel epitope tag that is absent in the mammalian proteome. Therefore, with a α MIN antibody, the MIN tag can be used for co-immunoprecipitation experiments, as well as microscopy.

So far, we characterized a GFP knockin construct, a BirA* cassette and a knockout cassette. The GFP knockin construct is useful to study spatio-temporal dynamics and kinetics of proteins in living cells. The BirA* cassette enables proximity-dependent labeling of the microenvironment (10 nm) of a given protein. This is mediated by the promiscuous biotin ligase activity of BirA* (Roux et al. 2012; D. I. Kim et al. 2014). The knockout cassette is used to induce a genetically defined gene disruption of any MIN tagged genes.

The above mentioned functional cassettes are by no means all possible functional modules. Inducible protein destruction/stabilization, expression or subcellular localization as well as enzymatic labeling of DNA binding sites by employing the *Escherichia coli* DNA adenine methyltransferase (DamID) would greatly advance the knowledge of protein function and protein-chromatin binding (van Steensel, Delrow, and Henikoff 2001; Banaszynski et al. 2006; Kennedy et al. 2010).

To validate the MIN tag strategy and its functional applications, we focused on a number of epigenetic factors that are well characterized and known in our laboratory and ESCs as a cellular model system. The MIN tag was successfully inserted in seven different genes: *DNMT1*, *DNMT3A*, *DNMT3B*, *TET1*, *TET2*, *TET3* and *UHRF1*.

First, we tested the applicability of the BirA* cassette, investigated the microenvironment of TET1 and could identify besides SIN3A other chromatin associated proteins including the closely related TET2. Costa *et al.* already could show that TET1 and TET2 have partially overlapping target sites (Costa et al. 2013), which is in accordance with our proteomics data. In conclusion, we could validate that the integration of the BirA* cassette into the locus of a gene of interest is a well-suited approach to investigate transient protein–protein interactions.

Second, we used the GFP knockin cassette to study subnuclear distribution of DNMT3B during differentiation. Early embryonic development is an interesting model to study a period of dramatic epigenetic change (see 1.3.2) that can be

recapitulated by differentiating naive ESCs to primed epiblast-like cells (EpiLCs) *in vitro* (Hayashi and Saitou 2013). The *de novo* DNA methyltransferase DNMT3B is important during the transition, which is accompanied by increased genome-wide methylation levels (Bachman 2001; T. Chen, Tsujimoto, and Li 2004; H. J. Lee, Hore, and Reik 2014), however the subnuclear distribution of DNMT3B during this dynamic process remains largely unknown. We could show that the global wave of *de novo* DNA methylation during epiblast differentiation follows a distinct spatio-temporal order, initiating at constitutive pericentromeric heterochromatin followed by facultative heterochromatin.

Exploiting the possibilities of the MIN tag strategy, we generated a cell line simultaneously expressing differentially tagged splicing isoforms of DNMT3B from different alleles. Interestingly, splicing isoforms of DNMT3B from different alleles behave differently. After 5-aza-dC treatment of ESCs, the major catalytically active isoform DNMT3B1 was immobilized at heterochromatic sites, while the DNMT3B6 fraction was not immobilized. This irreversible catalytic complex formation assay shows that only DNMT3B1 is catalytically active in this differentiation period.

The previously used integration cassettes were flanked by bacterial sequences derived from the original cloning constructs. It is known that foreign sequences such as bacterial sequences that are integrated into mammalian genomes become epigenetically silenced (Z. Y. Chen et al. 2004). To circumvent a potential silencing event occurring on recombined functional cassettes, a possibility could be to introduce solely eukaryotic expression cassettes that lack a bacterial backbone. A molecular approach could be the use of DNA molecules called minicircles (MCs) that solely contain a eukaryotic expression cassette as well as an integrase attachment site and lack a bacterial origin of replication as well as antibiotic resistance genes (Darquet et al. 1997, 1999).

Taken together, combination of the two genome engineering approaches, CRISPR/Cas assisted genomic sequence alterations and BXB1 mediated recombinations of functional cassettes, enables the generation of multiple cell lines with different functional derivatives from a single entry cell line. This approach can facilitate an easy and fast investigation of functional implications of mutations found in cancer and other human diseases.

3.4 Comparative Analysis of Single-Cell RNA Sequencing Methods

Over the last years, advancements in whole-transcriptome analyses have been achieved to investigate systemic biological questions. Most RNAseq analyses are performed on tissue samples or cell populations, in which biological differences between cells can be masked by averaging or mistaken for technical noise (Macaulay and Voet 2014). Single-cell approaches revolutionized the methodological approach to understand genomic, epigenomic and transcriptomic diversity that occurs during embryonic and even cancer development. Recent studies investigate cell-to-cell variability with scRNA-seq and thereby open up new possibilities to address highly complex biological questions, such as the development of heterogenous cell states within a tumor that is one of the major obstacles for cancer treatment (Ebinger et al. 2016; H. Li et al. 2017). To date, different high-throughput methods for scRNA-seq have been introduced that vary in coverage, sensitivity and accuracy.

In this study, I participated in comparing six prominent scRNA-seq protocols (Drop-seq, MARS-seq, SCRB-seq, Smart-seq2, Smart-seq/C1 and CEL-seq2/C1) with regard to the respective strengths and weaknesses in answering specific biological questions. The corresponding scRNA-seq libraries were generated in the same laboratory from wt mESCs, cultured under the same condition, which facilitates a comprehensive comparison across thousands of endogenous genes. To measure the sensitivity of each method, 96 different exogenous, poly(A)-tail positive, synthetic RNA transcripts were spiked-in with a known concentration developed by the External RNA Control Consortium (ERCC) (Jiang et al. 2011). The ERCC RNA standards undergo the steps of library preparation and thereby can be compared to the endogenous sample.

By comparing the different sensitivity estimates of the six methods, Drop-seq and MARS-seq were found to be the least sensitive, while Smart-seq2 is the most sensitive method, followed by SCRB-seq, Smart-seq/C1 and CEL-seq2/C1. For the detection of splicing events and different splicing forms an even read coverage across transcripts is necessary (Marinov et al. 2014). If an alternative splicing analysis is the focus, Smart-seq2 is the method of choice. Probably related to its higher sensitivity, it is also the most accurate method as the correlation of the ERCC spike-ins was the highest.

If the focus of the analysis lies on comparing the total RNA content of cells, known concentration differences of transcripts across cells would be required. In this regard, CEL-seq2/C1, MARS-seq, and SCRB-seq would have a strong advantage as UMIs and ERCCs can be used and compared.

Discussion

To compare relative expression across cells of any gene of interest, technical variation and the dropout probability are important factors to consider. Smart-seq2 performs best concerning dropout probability as it has a high sensitivity to detect even low expressed genes, while in UMI based methods perform better concerning technical variation as read duplicates that originate from PCR amplification can be excluded from down-stream analysis.

To assess the power to detect differentially expressed genes (Vieth et al. 2017), we combined the effects of dropout probability with technical variance and performed simulations for differential gene expression scenarios. In this analysis, SCRB-seq has a high sensitivity and a low amplification noise and also the highest power in scRNA-seq data at sequencing depth of one million reads, while in bulk RNA-seq libraries the effect of UMIs on power is minor (Parekh et al. 2016).

Besides the level of statistical power, also the expenses to generate the respective libraries and the sequencing costs per one million reads can be a decisive factor. Drop-seq is the most cost-effective method, but also generates a higher subset of reads that come from unfit cells and only has a low flexibility concerning the number of prepared libraries. SCRB-seq, MARS-seq, and Smart-seq2 follow Drop-seq concerning costs and are similarly cost-effective. Smart-seq/C1 and Smart-seq2 are the most expensive approaches. But using an in-house-purified Tn5 can drastically decrease the costs (Mora-Castilla et al. 2016).

In summary, the choice of the suitable method depends on the biological/experimental question. Drop-seq is the approach of choice when a rare cell type in a large cell pool is under investigation and the coverage is not essential for the downstream analysis. SCRB-seq and MARS-seq are preferable when quantifying transcriptomes of fewer cells, and Smart-seq2 is preferable when transcriptome annotation, identification of sequence variants, or the quantification of alternative splicing events are of interest, preferentially using the in-house-produced Tn5. ScRNA-seq protocols can still be improved by combining their strengths to become even more efficient. The results of our comparative analyses of six currently prominent scRNA-seq methods may facilitate such developments, and allow an informed choice among the tested methods.

3.5 Ubiquitome analysis reveals PAF as a specific ubiquitination target of UHRF1 in embryonic stem cells

E3 ligases mediate the final step of ubiquitin attachment to a target protein, thereby influencing protein localization, cell cycle progression, DNA repair and transcription (Hicke and Dunn 2003; Pickart 2004; L. Sun and Chen 2004; Pickart and Fushman 2004). The specific activity of most E3 ligases is mediated by a RING domain, which binds to an E2~ubiquitin thioester and initiates the transfer of the ubiquitin entity. E3 ligases harboring a RING domain are specified by more than 600 human genes, while protein kinases are encoded by 518 genes. Accordingly, RING E3 ligases have been linked to the control of many cellular processes and to multiple human diseases (Deshaies and Joazeiro 2009). Despite their critical importance, our knowledge of the physiological partners and biological functions for most RING E3 ligases remains limited.

Recently, we were able to show that binding of DNMT1 to ubiquitinated H3K18, which is set by the RING domain of UHRF1, is essential for DNA methylation (see 3.2). Prompted by our ubiquitination data (see 2.2), we set out to systematically screen also for non-histone targets of UHRF1 and its paralogue UHRF2, which has the same conserved motifs and overall structure as UHRF1.

In this study, we investigated specific ubiquitination targets of E3 ligases UHRF1 and UHRF2 in mouse ESCs by comparing isogenic *UHRF1* and *UHRF2* single KO ESC lines (see in 3.3) to exclude redundant ubiquitination targets. We used a published proteomics approach (Udeshi et al. 2013) to perform an unbiased, proteome-wide and site-specific analysis of ubiquitination levels. We enriched ubiquitin remnant motif-containing peptides (K-gly-gly) using an α -K-gly-gly antibody. This proteomics approach is a powerful method for unbiased quantification of former ubiquitinated peptides. However, this approach also has some limitations. It can only identify the sites of ubiquitination, but provides no information about the type of ubiquitin linkage. After trypsin digestion, the K-gly-gly motif can result from ubiquitin- but also NEDD8- or ISG15-modified proteins. However, K-gly-gly-containing peptides for the most part originate from ubiquitinated proteins in cultured cells (W. Kim et al. 2011). Despite the above mentioned limitations, this approach can provide information about sites of ubiquitination and further help to generate hypotheses, which can be then tested using biochemical and cell biological assays.

With this approach, we were able to identify several differentially ubiquitinated proteins that encompass biological processes such as transcriptional regulation, RNA binding, DNA damage response and cell cycle regulation. We find

Discussion

ubiquitination targets for both UHRF1 and UHRF2 such as HSP90, DNMT3B as well as UHRF1 specific targets such as UHRF1 itself, Trim28 and H3K18 (Citterio et al. 2004; Meilinger et al. 2009; Quenneville et al. 2011; Qin et al. 2015; Ding et al. 2016). Further, consistent with studies of UHRF1 activity *in vitro* and *in vivo*, we find differentially ubiquitinated histones: H3, H2A, H2B (Citterio et al. 2004; Harrison et al. 2016).

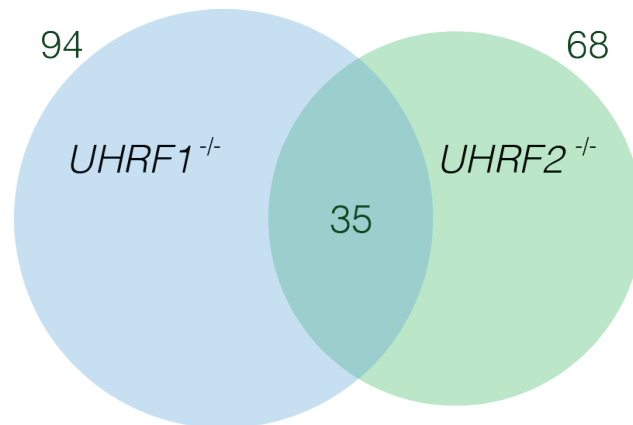


Figure 13. Overview of differentially ubiquitinated peptides. Venn diagram showing differentially ubiquitinated K-gly-gly peptides (gene name_position of ubiquitination) in *UHRF1* and *UHRF2* single knockout cells (Limma adj. p-value < 0.05). Modified from (Smets et al., 2017, manuscript in preparation).

Among the differentially ubiquitinated peptides, we focused on PCNA-associated factor (PAF) 15, the protein undergoing the highest loss of ubiquitination upon UHRF1 depletion. Mono-ubiquitination of PAF15 at Lys 15 and 24 is necessary for its interaction with PCNA and was described to be associated with translesion synthesis (TLS) inhibition during undisturbed S phase (Figure 14). Stalled replication caused by DNA damage sites leads to ubiquitin chain elongation on PAF15 by an unknown E3 ligase and subsequent degradation, which then allows for TLS polymerase recruitment to PCNA (Povlsen et al. 2012).

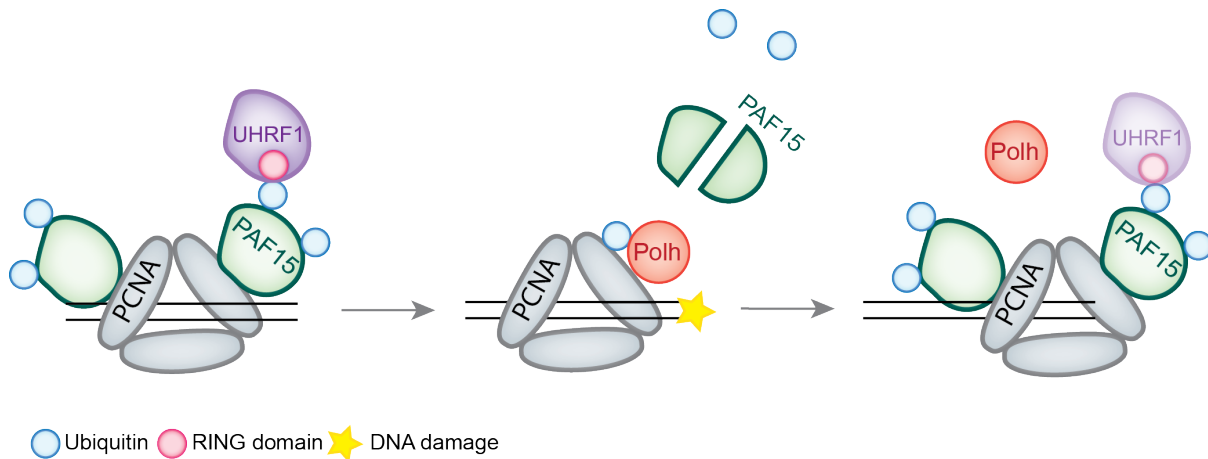


Figure 14. Model of the role of PAF15 in TLS. During undisturbed DNA replication, PCNA-bound PAF15 is ubiquitinated on Lys 15 and 24 by UHRF1. When PCNA encounters a site of DNA damage, ubiquitinated PAF15 is displaced from PCNA, which facilitates the interaction between TLS polymerases with mono-ubiquitinated PCNA (K164) at the stalled fork allowing lesion bypass. After the damage bypass, the reassociation of PAF15 with PCNA may help to promote reinitiation of normal replication. UHRF1 might also be responsible for ubiquitination of PAF15 after DNA damage bypass. Modified from (Povlsen et al. 2012; De Biasio et al. 2015).

However, the E3 ligase responsible for PAF15 double mono-ubiquitination remained unknown (Xie, Yao, and Dong 2014). Here, we show that the RING domain of UHRF1 ubiquitinates PAF15 at Lys 15 and 24 and influences its association with PCNA throughout S phase. As the interaction of PAF with PCNA is necessary for the ubiquitination step (Povlsen et al. 2012) and our high resolution microscopy analyses revealed co-localization with PCNA and UHRF1 exclusively during late S phase, we suggest that the ubiquitination takes place in a cell cycle dependent manner. Furthermore, both UHRF1 depletion and mutation of the lysine residues result in loss of PAF15 association with PCNA, which hints towards a role for PAF15 ubiquitination in stabilizing the PAF15-PCNA complex during replication. We speculate that the ubiquitination mark could potentially be recognized by proteins in close proximity that harbor ubiquitin interacting domains such as the UIM of DNMT1 (see 3.2). In summary, this study provides a novel aspect of UHRF1 in regulating replication block bypass by PAF15 ubiquitination. Our findings indicate an additional role of UHRF1 besides in DNA maintenance methylation but in replication-dependent DNA damage response.

Moreover and despite the sequence and domain similarity of UHRF1 and UHRF2, we also found specific ubiquitination targets of UHRF2 (Figure 13). Thus, a comprehensive follow-up research question will be whether and to what extent UHRF1 and UHRF2, two proteins with remarkable domain similarity, exhibit different

Discussion

biological functions and target distinct sets of proteins for ubiquitination. The current proteomic approach described in this thesis provides a resource for potential ubiquitination targets of UHRF2 for further investigation.

In summary, my findings suggest new functional aspects for UHRF1 both as a reader and writer of epigenetic histone modifications. In consequence, UHRF1 provides a possible link between different repressive signals, including H3K9 methylation and H3K18 ubiquitination.

Finally, proteome analysis showed that both UHRF1 and UHRF2 have distinct functions and targets beyond DNA methylation, arguing for non-redundant functions and possibly linking different cellular processes.

4 Annex

4.1 References

- Achour, M., X. Jacq, P. Rondé, M. Alhosin, C. Charlot, T. Chataigneau, M. Jeanblanc, et al. 2008. "The Interaction of the SRA Domain of ICBP90 with a Novel Domain of DNMT1 Is Involved in the Regulation of VEGF Gene Expression." *Oncogene* 27 (15): 2187–97.
- Allfrey, V. G., R. Faulkner, and A. E. Mirsky. 1964. "ACETYLATION AND METHYLATION OF HISTONES AND THEIR POSSIBLE ROLE IN THE REGULATION OF RNA SYNTHESIS." *Proceedings of the National Academy of Sciences of the United States of America* 51 (May): 786–94.
- Amouroux, Rachel, Buhe Nashun, Kenjiro Shirane, Shoma Nakagawa, Peter W. S. Hill, Zelpha D'Souza, Manabu Nakayama, et al. 2016. "De Novo DNA Methylation Drives 5hmC Accumulation in Mouse Zygotes." *Nature Cell Biology* 18 (2): 225–33.
- Anton, Tobias, and Sebastian Bultmann. 2017. "Site-Specific Recruitment of Epigenetic Factors with a Modular CRISPR/Cas System." *Nucleus*, February, 1–8.
- Aravin, Alexei A., Ravi Sachidanandam, Deborah Bourc'his, Christopher Schaefer, Dubravka Pezic, Katalin Fejes Toth, Timothy Bestor, and Gregory J. Hannon. 2008. "A piRNA Pathway Primed by Individual Transposons Is Linked to de Novo DNA Methylation in Mice." *Molecular Cell* 31 (6): 785–99.
- Arita, Kyohei, Mariko Ariyoshi, Hidehito Tochio, Yusuke Nakamura, and Masahiro Shirakawa. 2008. "Recognition of Hemi-Methylated DNA by the SRA Protein UHRF1 by a Base-Flipping Mechanism." *Nature* 455 (7214): 818–21.
- Avvakumov, George V., John R. Walker, Sheng Xue, Yanjun Li, Shili Duan, Christian Bronner, Cheryl H. Arrowsmith, and Sirano Dhe-Paganon. 2008. "Structural Basis for Recognition of Hemi-Methylated DNA by the SRA Domain of Human UHRF1." *Nature* 455 (7214): 822–25.
- Azuara, Véronique, Pascale Perry, Stephan Sauer, Mikhail Spivakov, Helle F. Jørgensen, Rosalind M. John, Mina Gouti, et al. 2006. "Chromatin Signatures of Pluripotent Cell Lines." *Nature Cell Biology* 8 (5): 532–38.
- Bachman, K. E. 2001. "Dnmt3a and Dnmt3b Are Transcriptional Repressors That Exhibit Unique Localization Properties to Heterochromatin." *The Journal of Biological Chemistry* 276 (34): 32282–87.
- Baets, Jonathan, Xiaohui Duan, Yanhong Wu, Gordon Smith, William W. Seeley, Inès Mademan, Nicole M. McGrath, et al. 2015. "Defects of Mutant DNMT1 Are Linked to a Spectrum of Neurological Disorders." *Brain: A Journal of Neurology* 138 (Pt 4): 845–61.
- Banaszynski, Laura A., Ling-Chun Chen, Lystranne A. Maynard-Smith, A. G. Lisa Ooi, and Thomas J. Wandless. 2006. "A Rapid, Reversible, and Tunable Method to Regulate Protein Function in Living Cells Using Synthetic Small Molecules." *Cell* 126 (5): 995–1004.
- Bannister, Andrew J., and Tony Kouzarides. 2011. "Regulation of Chromatin by Histone Modifications." *Cell Research* 21 (3): 381–95.
- Barau, Joan, Aurélie Teissandier, Natasha Zamudio, Stéphanie Roy, Valérie Nalesso, Yann Héroult, Florian Guillou, and Déborah Bourc'his. 2016. "The DNA Methyltransferase DNMT3C Protects Male Germ Cells from Transposon Activity." *Science* 354 (6314): 909–12.
- Barrera, Leah O., Zirong Li, Andrew D. Smith, Karen C. Arden, Webster K. Cavenee, Michael Q. Zhang, Roland D. Green, and Bing Ren. 2008. "Genome-Wide Mapping and Analysis of Active Promoters in Mouse Embryonic Stem Cells and Adult Organs." *Genome Research* 18 (1): 46–59.
- Barski, Artem, Suresh Cuddapah, Kairong Cui, Tae-Young Roh, Dustin E. Schones,

- Zhibin Wang, Gang Wei, Iouri Chepelev, and Keji Zhao. 2007. "High-Resolution Profiling of Histone Methylations in the Human Genome." *Cell* 129 (4): 823–37.
- Bashtrykov, Pavel, Gytis Jankevicius, Renata Z. Jurkowska, Sergey Ragozin, and Albert Jeltsch. 2014. "The UHRF1 Protein Stimulates the Activity and Specificity of the Maintenance DNA Methyltransferase DNMT1 by an Allosteric Mechanism." *The Journal of Biological Chemistry* 289 (7): 4106–15.
- Bashtrykov, Pavel, Arumugam Rajavelu, Benjamin Hackner, Sergey Ragozin, Thomas Carell, and Albert Jeltsch. 2014. "Targeted Mutagenesis Results in an Activation of DNA Methyltransferase 1 and Confirms an Autoinhibitory Role of Its RFTS Domain." *Chembiochem: A European Journal of Chemical Biology* 15 (5): 743–48.
- Batta, Kiran, Zhenhai Zhang, Kuangyu Yen, David B. Goffman, and B. Franklin Pugh. 2011. "Genome-Wide Function of H2B Ubiquitylation in Promoter and Genic Regions." *Genes & Development* 25 (21): 2254–65.
- Berkyurek, Ahmet Can, Isao Suetake, Kyohei Arita, Kohei Takeshita, Atsushi Nakagawa, Masahiro Shirakawa, and Shoji Tajima. 2014. "The DNA Methyltransferase Dnmt1 Directly Interacts with the SET and RING Finger-Associated (SRA) Domain of the Multifunctional Protein Uhrf1 to Facilitate Accession of the Catalytic Center to Hemi-Methylated DNA." *The Journal of Biological Chemistry* 289 (1): 379–86.
- Bernstein, Bradley E., Tarjei S. Mikkelsen, Xie Xiaohui, Kamal Michael, Dana J. Huebert, Cuff James, Fry Ben, et al. 2006. "A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells." *Cell* 125 (2): 315–26.
- Bestor, T. H., and V. M. Ingram. 1983. "Two DNA Methyltransferases from Murine Erythroleukemia Cells: Purification, Sequence Specificity, and Mode of Interaction with DNA." *Proceedings of the National Academy of Sciences of the United States of America* 80 (18): 5559–63.
- Bhaumik, Sukesh R., Edwin Smith, and Ali Shilatifard. 2007. "Covalent Modifications of Histones during Development and Disease Pathogenesis." *Nature Structural & Molecular Biology* 14 (11): 1008–16.
- Bird, Adrian. 2002. "DNA Methylation Patterns and Epigenetic Memory." *Genes & Development* 16 (1): 6–21.
- Blackledge, Neil P., Anca M. Farcas, Takashi Kondo, Hamish W. King, Joanna F. McGouran, Lars L. P. Hanssen, Shinsuke Ito, et al. 2014. "Variant PRC1 Complex-Dependent H2A Ubiquitylation Drives PRC2 Recruitment and Polycomb Domain Formation." *Cell* 157 (6): 1445–59.
- Borrelli, Emiliana, Eric J. Nestler, C. David Allis, and Paolo Sassone-Corsi. 2008. "Decoding the Epigenetic Language of Neuronal Plasticity." *Neuron* 60 (6): 961–74.
- Borsatti, Federica, and Mauro Mandrioli. 2004. "The Structure of Insect DNA Methyltransferase 2 (DNMT2) DNA Binding Domain Is Responsible for the Non-CpG Methylation in Insect Genomes." *Caryologia* 57 (3): 305–11.
- Bostick, Magnolia, Jong Kyong Kim, Pierre-Olivier Estève, Amander Clark, Sriharsa Pradhan, and Steven E. Jacobsen. 2007. "UHRF1 Plays a Role in Maintaining DNA Methylation in Mammalian Cells." *Science* 317 (5845): 1760–64.
- Bourc'his, Déborah, and Timothy H. Bestor. 2004. "Meiotic Catastrophe and Retrotransposon Reactivation in Male Germ Cells Lacking Dnmt3L." *Nature* 431 (7004): 96–99.
- Bourc'his, D., P. Miniou, M. Jeanpierre, D. Molina Gomes, J-M Dupont, G. De Saint-Basile, P. Maraschio, L. Tiepolo, and E. Viegas-Péquignot. 1999. "Abnormal Methylation Does Not Prevent X Inactivation in ICF Patients." *Cytogenetic and Genome Research* 84 (3-4): 245–52.
- Braun, S. M. G., and S. Jessberger. 2014. "Review: Adult Neurogenesis and Its Role in Neuropsychiatric Disease, Brain Repair and Normal Brain Function." *Neuropathology and Applied Neurobiology* 40 (1): 3–12.
- Bronner, Christian, Mayada Achour, Yoshimi Arima, Thierry Chataigneau, Hideyuki Saya, and Valérie B. Schini-Kerth. 2007. "The UHRF Family: Oncogenes That

- Are Drugable Targets for Cancer Therapy in the near Future?" *Pharmacology & Therapeutics* 115 (3): 419–34.
- Brooks, P. J., C. Marietta, and D. Goldman. 1996. "DNA Mismatch Repair and DNA Methylation in Adult Brain Neurons." *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 16 (3): 939–45.
- Brown, William R. A., Nicholas C. O. Lee, Zhengyao Xu, and Margaret C. M. Smith. 2011. "Serine Recombinases as Tools for Genome Engineering." *Methods* 53 (4): 372–79.
- Capecchi, Mario R. 2005. "Gene Targeting in Mice: Functional Analysis of the Mammalian Genome for the Twenty-First Century." *Nature Reviews. Genetics* 6 (6): 507–12.
- Cardoso, M. C., and H. Leonhardt. 1999. "DNA Methyltransferase Is Actively Retained in the Cytoplasm during Early Development." *The Journal of Cell Biology* 147 (1): 25–32.
- Chalkiadaki, Angeliki, and Leonard Guarente. 2015. "The Multifaceted Functions of Sirtuins in Cancer." *Nature Reviews. Cancer* 15 (10): 608–24.
- Chedin, Frederic, Michael R. Lieber, and Chih-Lin Hsieh. 2002. "The DNA Methyltransferase-like Protein DNMT3L Stimulates de Novo Methylation by Dnmt3a." *Proceedings of the National Academy of Sciences of the United States of America* 99 (26): 16916–21.
- Cheng, Jingdong, Yi Yang, Jian Fang, Jianxiong Xiao, Tingting Zhu, Fei Chen, Ping Wang, Ze Li, Huirong Yang, and Yanhui Xu. 2013. "Structural Insight into Coordinated Recognition of Trimethylated Histone H3 Lysine 9 (H3K9me3) by the Plant Homeodomain (PHD) and Tandem Tudor Domain (TTD) of UHRF1 (ubiquitin-Like, Containing PHD and RING Finger Domains, 1) Protein." *The Journal of Biological Chemistry* 288 (2): 1329–39.
- Cheng, Xiaodong, and Robert M. Blumenthal. 2008. "Mammalian DNA Methyltransferases: A Structural Perspective." *Structure* 16 (3): 341–50.
- Chen, Ruoyu, Qiao Zhang, Xiaoya Duan, Philippe York, Guo-Dong Chen, Pengcheng Yin, Haijun Zhu, et al. 2017. "The 5-Hydroxymethylcytosine (5hmC) Reader UHRF2 Is Required for Normal Levels of 5hmC in Mouse Adult Brain and Spatial Learning and Memory." *The Journal of Biological Chemistry* 292 (11): 4533–43.
- Chen, T., N. Tsujimoto, and E. Li. 2004. "The PWWP Domain of Dnmt3a and Dnmt3b Is Required for Directing DNA Methylation to the Major Satellite Repeats at Pericentric Heterochromatin." *Molecular and Cellular Biology* 24 (20): 9048–58.
- Chen, Z. Y., C. Y. He, L. Meuse, and M. A. Kay. 2004. "Silencing of Episomal Transgene Expression by Plasmid Bacterial DNA Elements in Vivo." *Gene Therapy* 11 (10): 856–64.
- Choi, Jae Duk, and Lee Jong-Soo. 2013. "Interplay between Epigenetics and Genetics in Cancer." *Genomics & Informatics* 11 (4): 164.
- Chuang, L. S., H. I. Ian, T. W. Koh, H. H. Ng, G. Xu, and B. F. Li. 1997. "Human DNA-(cytosine-5) Methyltransferase-PCNA Complex as a Target for p21WAF1." *Science* 277 (5334): 1996–2000.
- Citterio, Elisabetta, Roberto Papait, Francesco Nicassio, Manuela Vecchi, Paola Gomiero, Roberto Mantovani, Pier Paolo Di Fiore, and Ian Marc Bonapace. 2004. "Np95 Is a Histone-Binding Protein Endowed with Ubiquitin Ligase Activity." *Molecular and Cellular Biology* 24 (6): 2526–35.
- Cooper, Sarah, Martin Dienstbier, Raihann Hassan, Lothar Schermelleh, Jafar Sharif, Neil P. Blackledge, Valeria De Marco, et al. 2014. "Targeting Polycomb to Pericentric Heterochromatin in Embryonic Stem Cells Reveals a Role for H2AK119u1 in PRC2 Recruitment." *Cell Reports* 7 (5): 1456–70.
- Costa, Yael, Junjun Ding, Thorold W. Theunissen, Francesco Faiola, Timothy A. Hore, Pavel V. Shliha, Miguel Fidalgo, et al. 2013. "NANOG-Dependent Function of TET1 and TET2 in Establishment of Pluripotency." *Nature* 495 (7441): 370–74.
- Creyghton, M. P., A. W. Cheng, G. G. Welstead, T. Kooistra, B. W. Carey, E. J. Steine, J. Hanna, et al. 2010. "Histone H3K27ac Separates Active from Poised

- Enhancers and Predicts Developmental State." *Proceedings of the National Academy of Sciences* 107 (50): 21931–36.
- Cuthbert, Graeme L., Sylvain Daujat, Andrew W. Snowden, Hediye Erdjument-Bromage, Teruki Hagiwara, Michiyuki Yamada, Robert Schneider, et al. 2004. "Histone Deimination Antagonizes Arginine Methylation." *Cell* 118 (5): 545–53.
- Danzer, John R., and Lori L. Wallrath. 2004. "Mechanisms of HP1-Mediated Gene Silencing in *Drosophila*." *Development* 131 (15): 3571–80.
- Darquet, A. M., B. Cameron, P. Wils, D. Scherman, and J. Crouzet. 1997. "A New DNA Vehicle for Nonviral Gene Delivery: Supercoiled Minicircle." *Gene Therapy* 4 (12): 1341–49.
- Darquet, A. M., R. Rangara, P. Kreiss, B. Schwartz, S. Naimi, P. Delaère, J. Crouzet, and D. Scherman. 1999. "Minicircle: An Improved DNA Molecule for in Vitro and in Vivo Gene Transfer." *Gene Therapy* 6 (2): 209–18.
- David, Gosal, Ealing John, and Mignot Emmanuel. 2013. "A MUTATION IN THE DNMT1 GENE CAUSING AUTOSOMAL DOMINANT ATAXIA WITH DEAFNESS AND CATAPLEXY." *Journal of Neurology, Neurosurgery, and Psychiatry* 84 (11). BMJ Publishing Group Ltd: e2–e2.
- Dawlaty, Meelad M., Achim Breiling, Thuc Le, Günter Raddatz, M. Inmaculada Barrasa, Albert W. Cheng, Qing Gao, et al. 2013. "Combined Deficiency of Tet1 and Tet2 Causes Epigenetic Abnormalities but Is Compatible with Postnatal Development." *Developmental Cell* 24 (3): 310–23.
- Dawlaty, Meelad M., Kibibi Ganz, Benjamin E. Powell, Yueh-Chiang Hu, Styliani Markoulaki, Albert W. Cheng, Qing Gao, et al. 2011. "Tet1 Is Dispensable for Maintaining Pluripotency and Its Loss Is Compatible with Embryonic and Postnatal Development." *Cell Stem Cell* 9 (2): 166–75.
- Deaton, A. M., and A. Bird. 2011. "CpG Islands and the Regulation of Transcription." *Genes & Development* 25 (10): 1010–22.
- De Biasio, Alfredo, Alain Ibáñez de Opakua, Gulnazar B. Mortuza, Rafael Molina, Tiago N. Cordeiro, Francisco Castillo, Maider Villate, et al. 2015. "Structure of p15PAF–PCNA Complex and Implications for Clamp Sliding during DNA Replication and Repair." *Nature Communications* 6 (March). Nature Publishing Group: 6439.
- Deckert, J., and K. Struhl. 2001. "Histone Acetylation at Promoters Is Differentially Affected by Specific Activators and Repressors." *Molecular and Cellular Biology* 21 (8): 2726–35.
- Deshaies, Raymond J., and Claudio A. P. Joazeiro. 2009. "RING Domain E3 Ubiquitin Ligases." *Annual Review of Biochemistry* 78: 399–434.
- Dhayalan, Arunkumar, Arumugam Rajavelu, Philipp Rathert, Raluca Tamas, Renata Z. Jurkowska, Sergey Ragozin, and Albert Jeltsch. 2010. "The Dnmt3a PWWP Domain Reads Histone 3 Lysine 36 Trimethylation and Guides DNA Methylation." *The Journal of Biological Chemistry* 285 (34): 26114–20.
- Ding, Guangjin, Peilin Chen, Hui Zhang, Xiaojie Huang, Yi Zang, Jiwen Li, Jia Li, and Jiemin Wong. 2016. "Regulation of Ubiquitin-like with Plant Homeodomain and RING Finger Domain 1 (UHRF1) Protein Stability by Heat Shock Protein 90 Chaperone Machinery." *The Journal of Biological Chemistry* 291 (38): 20125–35.
- Di Ruscio, Annalisa, Alexander K. Ebralidze, Touati Benoukraf, Giovanni Amabile, Loyal A. Goff, Jolyon Terragni, Maria Eugenia Figueroa, et al. 2013. "DNMT1-Interacting RNAs Block Gene-Specific DNA Methylation." *Nature* 503 (7476): 371–76.
- Du, Zhanwen, Jing Song, Yong Wang, Yiqing Zhao, Kishore Guda, Shuming Yang, Hung-Ying Kao, et al. 2010. "DNMT1 Stability Is Regulated by Proteins Coordinating Deubiquitination and Acetylation-Driven Ubiquitination." *Science Signaling* 3 (146): ra80.
- Easwaran, Hariharan P., Lothar Schermelleh, Heinrich Leonhardt, and M. Cristina Cardoso. 2004. "Replication-Independent Chromatin Loading of Dnmt1 during G2 and M Phases." *EMBO Reports* 5 (12): 1181–86.
- Ebinger, Sarah, Erbey Ziya Özdemir, Christoph Ziegenhain, Sebastian Tiedt,

- Catarina Castro Alves, Michaela Grunert, Michael Dworzak, et al. 2016. "Characterization of Rare, Dormant, and Therapy-Resistant Cells in Acute Lymphoblastic Leukemia." *Cancer Cell*, November. doi:10.1016/j.ccell.2016.11.002.
- Ehrlich, M., M. A. Gama-Sosa, L. H. Huang, R. M. Midgett, K. C. Kuo, R. A. McCune, and C. Gehrke. 1982. "Amount and Distribution of 5-Methylcytosine in Human DNA from Different Types of Tissues of Cells." *Nucleic Acids Research* 10 (8): 2709–21.
- Estève, Pierre-Olivier, Yanqi Chang, Mala Samaranayake, Anup K. Upadhyay, John R. Horton, George R. Feehery, Xiaodong Cheng, and Sriharsa Pradhan. 2011. "A Methylation and Phosphorylation Switch between an Adjacent Lysine and Serine Determines Human DNMT1 Stability." *Nature Structural & Molecular Biology* 18 (1): 42–48.
- Fan, G., C. Beard, R. Z. Chen, G. Csankovszki, Y. Sun, M. Siniaia, D. Biniszkievicz, et al. 2001. "DNA Hypomethylation Perturbs the Function and Survival of CNS Neurons in Postnatal Animals." *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 21 (3): 788–97.
- Fatemi, M., A. Hermann, S. Pradhan, and A. Jeltsch. 2001. "The Activity of the Murine DNA Methyltransferase Dnmt1 Is Controlled by Interaction of the Catalytic Domain with the N-Terminal Part of the Enzyme Leading to an Allosteric Activation of the Enzyme after Binding to Methylated DNA." *Journal of Molecular Biology* 309 (5): 1189–99.
- Feldman, Nirit, Ariela Gerson, Jia Fang, En Li, Yi Zhang, Yoichi Shinkai, Howard Cedar, and Yehudit Bergman. 2006. "G9a-Mediated Irreversible Epigenetic Inactivation of Oct-3/4 during Early Embryogenesis." *Nature Cell Biology* 8 (2): 188–94.
- Felle, Max, Saskia Joppien, Attila Németh, Sarah Diermeier, Verena Thalhammer, Thomas Dobner, Elisabeth Kremmer, Roland Kappler, and Gernot Längst. 2011. "The USP7/Dnmt1 Complex Stimulates the DNA Methylation Activity of Dnmt1 and Regulates the Stability of UHRF1." *Nucleic Acids Research* 39 (19): 8355–65.
- Feng, Jian, Hua Chang, En Li, and Guoping Fan. 2005. "Dynamic Expression of de Novo DNA Methyltransferases Dnmt3a and Dnmt3b in the Central Nervous System." *Journal of Neuroscience Research* 79 (6): 734–46.
- Ferrari, Karin J., Andrea Scelfo, Sriganesh Jammula, Alessandro Cuomo, Iros Barozzi, Alexandra Stützer, Wolfgang Fischle, Tiziana Bonaldi, and Diego Pasini. 2014. "Polycomb-Dependent H3K27me1 and H3K27me2 Regulate Active Transcription and Enhancer Fidelity." *Molecular Cell* 53 (1): 49–62.
- Ficz, Gabriella, Miguel R. Branco, Stefanie Seisenberger, Fátima Santos, Felix Krueger, Timothy A. Hore, C. Joana Marques, Simon Andrews, and Wolf Reik. 2011. "Dynamic Regulation of 5-Hydroxymethylcytosine in Mouse ES Cells and during Differentiation." *Nature* 473 (7347): 398–402.
- Frauer, Carina, and Heinrich Leonhardt. 2009. "A Versatile Non-Radioactive Assay for DNA Methyltransferase Activity and DNA Binding." *Nucleic Acids Research* 37 (3): e22.
- Frauer, Carina, Andrea Rottach, Daniela Meilinger, Sebastian Bultmann, Karin Fellingner, Stefan Hasenöder, Mengxi Wang, et al. 2011. "Different Binding Properties and Function of CXXC Zinc Finger Domains in Dnmt1 and Tet1." *PLoS One* 6 (2): e16627.
- Fujimori, A., Y. Matsuda, Y. Takemoto, Y. Hashimoto, E. Kubo, R. Araki, R. Fukumura, K. Mita, K. Tatsumi, and M. Muto. 1998. "Cloning and Mapping of Np95 Gene Which Encodes a Novel Nuclear Protein Associated with Cell Proliferation." *Mammalian Genome: Official Journal of the International Mammalian Genome Society* 9 (12): 1032–35.
- Fuks, François, Wendy A. Burgers, Alexander Brehm, Luke Hughes-Davies, and Tony Kouzarides. 2000. "DNA Methyltransferase Dnmt1 Associates with Histone Deacetylase Activity." *Nature Genetics* 24 (1). Nature Publishing Group: 88–91.
- Fuks, François, Wendy A. Burgers, Nadia Godin, Masataka Kasai, and Tony

- Kouzarides. 2001. "Dnmt3a Binds Deacetylases and Is Recruited by a Sequence-specific Repressor to Silence Transcription." *The EMBO Journal* 20 (10): EMBO Press: 2536–44.
- Fuks, François, Paul J. Hurd, Rachel Deplus, and Tony Kouzarides. 2003. "The DNA Methyltransferases Associate with HP1 and the SUV39H1 Histone Methyltransferase." *Nucleic Acids Research* 31 (9): 2305–12.
- Gaudet, François, J. Graeme Hodgson, Amir Eden, Laurie Jackson-Grusby, Jessica Dausman, Joe W. Gray, Heinrich Leonhardt, and Rudolf Jaenisch. 2003. "Induction of Tumors in Mice by Genomic Hypomethylation." *Science* 300 (5618): 489–92.
- Gelato, Kathy A., Maria Tauber, Michelle S. Ong, Stefan Winter, Kyoko Hiragami-Hamada, Julia Sindlinger, Alexander Lemak, et al. 2014. "Accessibility of Different Histone H3-Binding Domains of UHRF1 Is Allosterically Regulated by Phosphatidylinositol 5-Phosphate." *Molecular Cell* 54 (6): 905–19.
- Ge, Ying-Zi, Min-Tie Pu, Humaira Gowher, Hai-Ping Wu, Jian-Ping Ding, Albert Jeltsch, and Guo-Liang Xu. 2004. "Chromatin Targeting of de Novo DNA Methyltransferases by the PWWP Domain." *The Journal of Biological Chemistry* 279 (24): 25447–54.
- Gjerset, R. A., and D. W. Martin Jr. 1982. "Presence of a DNA Demethylating Activity in the Nucleus of Murine Erythroleukemic Cells." *The Journal of Biological Chemistry* 257 (15): 8581–83.
- Globisch, Daniel, Martin Münzel, Markus Müller, Stylianos Michalakis, Mirko Wagner, Susanne Koch, Tobias Brückl, Martin Biel, and Thomas Carell. 2010. "Tissue Distribution of 5-Hydroxymethylcytosine and Search for Active Demethylation Intermediates." *PloS One* 5 (12): e15367.
- Goldberg, Aaron D., C. David Allis, and Emily Bernstein. 2007. "Epigenetics: A Landscape Takes Shape." *Cell* 128 (4): 635–38.
- Goll, Mary Grace, and Timothy H. Bestor. 2005. "Eukaryotic Cytosine Methyltransferases." *Annual Review of Biochemistry* 74: 481–514.
- Goll, Mary Grace, Finn Kirpekar, Keith A. Maggert, Jeffrey A. Yoder, Chih-Lin Hsieh, Xiaoyu Zhang, Kent G. Golic, Steven E. Jacobsen, and Timothy H. Bestor. 2006. "Methylation of tRNA^{Asp} by the DNA Methyltransferase Homolog Dnmt2." *Science* 311 (5759): 395–98.
- Goto, K., M. Numata, J. I. Komura, T. Ono, T. H. Bestor, and H. Kondo. 1994. "Expression of DNA Methyltransferase Gene in Mature and Immature Neurons as Well as Proliferating Cells in Mice." *Differentiation; Research in Biological Diversity* 56 (1-2): 39–44.
- Gowher, Humaira, Kirsten Liebert, Andrea Hermann, Guoliang Xu, and Albert Jeltsch. 2005. "Mechanism of Stimulation of Catalytic Activity of Dnmt3A and Dnmt3B DNA-(cytosine-C5)-Methyltransferases by Dnmt3L." *The Journal of Biological Chemistry* 280 (14): 13341–48.
- Guenther, Matthew G., Stuart S. Levine, Laurie A. Boyer, Jaenisch Rudolf, and Richard A. Young. 2007. "A Chromatin Landmark and Transcription Initiation at Most Promoters in Human Cells." *Cell* 130 (1): 77–88.
- Gu, Tian-Peng, Fan Guo, Hui Yang, Hai-Ping Wu, Gui-Fang Xu, Wei Liu, Zhi-Guo Xie, et al. 2011. "The Role of Tet3 DNA Dioxygenase in Epigenetic Reprogramming by Oocytes." *Nature* 477 (7366): 606–10.
- Hahn, Maria A., Runxiang Qiu, Xiwei Wu, Arthur X. Li, Heying Zhang, Jun Wang, Jonathan Jui, et al. 2013. "Dynamics of 5-Hydroxymethylcytosine and Chromatin Marks in Mammalian Neurogenesis." *Cell Reports* 3 (2): 291–300.
- Hajkova, Petra, Sean J. Jeffries, Caroline Lee, Nigel Miller, Stephen P. Jackson, and M. Azim Surani. 2010. "Genome-Wide Reprogramming in the Mouse Germ Line Entails the Base Excision Repair Pathway." *Science* 329 (5987): 78–82.
- Hansen, R. S., C. Wijmenga, P. Luo, A. M. Stanek, T. K. Canfield, C. M. Weemaes, and S. M. Gartler. 1999. "The DNMT3B DNA Methyltransferase Gene Is Mutated in the ICF Immunodeficiency Syndrome." *Proceedings of the National Academy of Sciences of the United States of America* 96 (25): 14412–17.
- Harrison, Joseph S., Evan M. Cornett, Dennis Goldfarb, Paul A. DaRosa, Zimeng M.

- Li, Feng Yan, Bradley M. Dickson, et al. 2016. "Hemi-Methylated DNA Regulates DNA Methylation Inheritance through Allosteric Activation of H3 Ubiquitylation by UHRF1." *eLife* 5 (September). doi:10.7554/eLife.17101.
- Hashimoto, Hideharu, John R. Horton, Xing Zhang, Magnolia Bostick, Steven E. Jacobsen, and Xiaodong Cheng. 2008. "The SRA Domain of UHRF1 Flips 5-Methylcytosine out of the DNA Helix." *Nature* 455 (7214): 826–29.
- Hashimshony, Tamar, Florian Wagner, Noa Sher, and Itai Yanai. 2012. "CEL-Seq: Single-Cell RNA-Seq by Multiplexed Linear Amplification." *Cell Reports* 2 (3): 666–73.
- Haurwitz, Rachel E., Martin Jinek, Blake Wiedenheft, Kaihong Zhou, and Jennifer A. Doudna. 2010. "Sequence- and Structure-Specific RNA Processing by a CRISPR Endonuclease." *Science* 329 (5997): 1355–58.
- Hayashi, Katsuhiko, and Mitunori Saitou. 2013. "Generation of Eggs from Mouse Embryonic Stem Cells and Induced Pluripotent Stem Cells." *Nature Protocols* 8 (8): 1513–24.
- Heo, Jinbeom, Jisun Lim, Seungun Lee, Jaeho Jeong, Hyunsook Kang, Yonghwan Kim, Jeong Wook Kang, et al. 2017. "Sirt1 Regulates DNA Methylation and Differentiation Potential of Embryonic Stem Cells by Antagonizing Dnmt3l." *Cell Reports* 18 (8): 1930–45.
- Hermann, A., S. Schmitt, and A. Jeltsch. 2003. "The Human Dnmt2 Has Residual DNA-(Cytosine-C5) Methyltransferase Activity." *The Journal of Biological Chemistry* 278 (34): 31717–21.
- He, Yu-Fei, Bin-Zhong Li, Zheng Li, Peng Liu, Yang Wang, Qingyu Tang, Jianping Ding, et al. 2011. "Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA." *Science* 333 (6047): 1303–7.
- Hicke, Linda, and Rebecca Dunn. 2003. "Regulation of Membrane Protein Transport by Ubiquitin and Ubiquitin-Binding Proteins." *Annual Review of Cell and Developmental Biology* 19 (1): 141–72.
- Hockemeyer, Dirk, Haoyi Wang, Samira Kiani, Christine S. Lai, Qing Gao, John P. Cassady, Gregory J. Cost, et al. 2011. "Genetic Engineering of Human Pluripotent Cells Using TALE Nucleases." *Nature Biotechnology* 29 (8): 731–34.
- Howlett, S. K., and W. Reik. 1991. "Methylation Levels of Maternal and Paternal Genomes during Preimplantation Development." *Development* 113 (1): 119–27.
- Inano, K., I. Suetake, T. Ueda, Y. Miyake, M. Nakamura, M. Okada, and S. Tajima. 2000. "Maintenance-Type DNA Methyltransferase Is Highly Expressed in Post-Mitotic Neurons and Localized in the Cytoplasmic Compartment." *Journal of Biochemistry* 128 (2): 315–21.
- Inoue, Azusa, and Yi Zhang. 2011. "Replication-Dependent Loss of 5-Hydroxymethylcytosine in Mouse Preimplantation Embryos." *Science* 334 (6053): 194.
- Iqbal, Khursheed, Seung-Gi Jin, Gerd P. Pfeifer, and Pirooska E. Szabó. 2011. "Reprogramming of the Paternal Genome upon Fertilization Involves Genome-Wide Oxidation of 5-Methylcytosine." *Proceedings of the National Academy of Sciences of the United States of America* 108 (9): 3642–47.
- Ito, Shinsuke, Ana C. D'Alessio, Olena V. Taranova, Kwonho Hong, Lawrence C. Sowers, and Yi Zhang. 2010. "Role of Tet Proteins in 5mC to 5hmC Conversion, ES-Cell Self-Renewal and Inner Cell Mass Specification." *Nature* 466 (7310): 1129–33.
- Ito, Shinsuke, Li Shen, Qing Dai, Susan C. Wu, Leonard B. Collins, James A. Swenberg, Chuan He, and Yi Zhang. 2011. "Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine." *Science* 333 (6047): 1300–1303.
- Jacobson, R. H., A. G. Ladurner, D. S. King, and R. Tjian. 2000. "Structure and Function of a Human TAFII250 Double Bromodomain Module." *Science* 288 (5470): 1422–25.
- Jaenisch, Rudolf, and Adrian Bird. 2003. "Epigenetic Regulation of Gene Expression: How the Genome Integrates Intrinsic and Environmental Signals." *Nature Genetics* 33 Suppl (March): 245–54.

- Jeltsch, Albert. 2006. "On the Enzymatic Properties of Dnmt1: Specificity, Processivity, Mechanism of Linear Diffusion and Allosteric Regulation of the Enzyme." *Epigenetics: Official Journal of the DNA Methylation Society* 1 (2): 63–66.
- Jeong, Shinwu, Gangning Liang, Shikhar Sharma, Joy C. Lin, Si Ho Choi, Han Han, Christine B. Yoo, Gerda Egger, Allen S. Yang, and Peter A. Jones. 2009. "Selective Anchoring of DNA Methyltransferases 3A and 3B to Nucleosomes Containing Methylated DNA." *Molecular and Cellular Biology* 29 (19): 5366–76.
- Jiang, Lichun, Felix Schlesinger, Carrie A. Davis, Yu Zhang, Renhua Li, Marc Salit, Thomas R. Gingeras, and Brian Oliver. 2011. "Synthetic Spike-in Standards for RNA-Seq Experiments." *Genome Research* 21 (9): 1543–51.
- Jia, Yuanhui, Pishun Li, Lan Fang, Haijun Zhu, Liangliang Xu, Hao Cheng, Junying Zhang, et al. 2016. "Negative Regulation of DNMT3A de Novo DNA Methylation by Frequently Overexpressed UHRF Family Proteins as a Mechanism for Widespread DNA Hypomethylation in Cancer." *Cell Discovery* 2 (April): 16007.
- Jung, Yong Wook, Eriona Hysolli, Kun-Yong Kim, Yoshiaki Tanaka, and In-Hyun Park. 2012. "Human Induced Pluripotent Stem Cells and Neurodegenerative Disease: Prospects for Novel Therapies." *Current Opinion in Neurology* 25 (2): 125–30.
- Kalb, Reinhard, Sebastian Latwiel, H. Irem Baymaz, Pascal W. T. C. Jansen, Christoph W. Müller, Michiel Vermeulen, and Jürg Müller. 2014. "Histone H2A Monoubiquitination Promotes Histone H3 Methylation in Polycomb Repression." *Nature Structural & Molecular Biology* 21 (6): 569–71.
- Kaneda, Masahiro, Masaki Okano, Kenichiro Hata, Takashi Sado, Naomi Tsujimoto, En Li, and Hiroyuki Sasaki. 2004. "Essential Role for de Novo DNA Methyltransferase Dnmt3a in Paternal and Maternal Imprinting." *Nature* 429 (6994): 900–903.
- Karagianni, Panagiota, Larbi Amazit, Jun Qin, and Jiemin Wong. 2008. "ICBP90, a Novel Methyl K9 H3 Binding Protein Linking Protein Ubiquitination with Heterochromatin Formation." *Molecular and Cellular Biology* 28 (2): 705–17.
- Kennedy, Matthew J., Robert M. Hughes, Leslie A. Peteya, Joel W. Schwartz, Michael D. Ehlers, and Chandra L. Tucker. 2010. "Rapid Blue-Light-Mediated Induction of Protein Interactions in Living Cells." *Nature Methods* 7 (12): 973–75.
- Keravala, Annahita, Amy C. Groth, Sohail Jarrahan, Bhaskar Thyagarajan, Jason J. Hoyt, Patrick J. Kirby, and Michele P. Calos. 2006. "A Diversity of Serine Phage Integrases Mediate Site-Specific Recombination in Mammalian Cells." *Molecular Genetics and Genomics: MGG* 276 (2): 135–46.
- Kim, Dae In, K. C. Birendra, Wenhong Zhu, Khatereh Motamedchaboki, Valérie Doye, and Kyle J. Roux. 2014. "Probing Nuclear Pore Complex Architecture with Proximity-Dependent Biotinylation." *Proceedings of the National Academy of Sciences of the United States of America* 111 (24): E2453–61.
- Kim, Jaehoon, Sandra B. Hake, and Robert G. Roeder. 2005. "The Human Homolog of Yeast BRE1 Functions as a Transcriptional Coactivator through Direct Activator Interactions." *Molecular Cell* 20 (5): 759–70.
- Kim, Jong Kyong, Pierre-Olivier Estève, Steven E. Jacobsen, and Sriharsa Pradhan. 2009. "UHRF1 Binds G9a and Participates in p21 Transcriptional Regulation in Mammalian Cells." *Nucleic Acids Research* 37 (2): 493–505.
- Kim, Woong, Eric J. Bennett, Edward L. Huttlin, Ailan Guo, Jing Li, Anthony Possemato, Mathew E. Sowa, et al. 2011. "Systematic and Quantitative Assessment of the Ubiquitin-Modified Proteome." *Molecular Cell* 44 (2): 325–40.
- Klein, Christopher J., Tom Bird, Nilufer Ertekin-Taner, Sarah Lincoln, Robert Hjorth, Yanhong Wu, John Kwok, Georges Mer, Peter J. Dyck, and Garth A. Nicholson. 2013. "DNMT1 Mutation Hot Spot Causes Varied Phenotypes of HSAN1 with Dementia and Hearing Loss." *Neurology* 80 (9): 824–28.
- Klein, Christopher J., Maria-Victoria Botuyan, Yanhong Wu, Christopher J. Ward, Garth A. Nicholson, Simon Hammans, Kaori Hojo, et al. 2011. "Mutations in DNMT1 Cause Hereditary Sensory Neuropathy with Dementia and Hearing

- Loss." *Nature Genetics* 43 (6): 595–600.
- Koh, Kian Peng, Akiko Yabuuchi, Sridhar Rao, Yun Huang, Kerrianne Cuniff, Julie Nardone, Asta Laiho, et al. 2011. "Tet1 and Tet2 Regulate 5-Hydroxymethylcytosine Production and Cell Lineage Specification in Mouse Embryonic Stem Cells." *Cell Stem Cell* 8 (2): 200–213.
- Kolodziejczyk, Aleksandra A., Jong Kyoung Kim, Jason C. H. Tsang, Tomislav Ilicic, Johan Henriksson, Kedar N. Natarajan, Alex C. Tuck, et al. 2015. "Single Cell RNA-Sequencing of Pluripotent States Unlocks Modular Transcriptional Variation." *Cell Stem Cell* 17 (4): 471–85.
- Kornberg, R. D., and Y. Lorch. 1999. "Twenty-Five Years of the Nucleosome, Fundamental Particle of the Eukaryote Chromosome." *Cell* 98 (3): 285–94.
- Kouzarides, Tony. 2007. "Chromatin Modifications and Their Function." *Cell* 128 (4): 693–705.
- Kriaucionis, Skirmantas, and Nathaniel Heintz. 2009. "The Nuclear DNA Base 5-Hydroxymethylcytosine Is Present in Purkinje Neurons and the Brain." *Science* 324 (5929): 929–30.
- Ku, Manching, Richard P. Koche, Esther Rheinbay, Eric M. Mendenhall, Mitsuhiro Endoh, Tarjei S. Mikkelsen, Aviva Presser, et al. 2008. "Genomewide Analysis of PRC1 and PRC2 Occupancy Identifies Two Classes of Bivalent Domains." *PLoS Genetics* 4 (10): e1000242.
- Kunert, Natascha, Joachim Marhold, Jonas Stanke, Dirk Stach, and Frank Lyko. 2003. "A Dnmt2-like Protein Mediates DNA Methylation in *Drosophila*." *Development* 130 (21): 5083–90.
- Kurdistani, Siavash K., Saeed Tavazoie, and Michael Grunstein. 2004. "Mapping Global Histone Acetylation Patterns to Gene Expression." *Cell* 117 (6): 721–33.
- Kuscu, Cem, Sevki Arslan, Ritambhara Singh, Jeremy Thorpe, and Mazhar Adli. 2014. "Genome-Wide Analysis Reveals Characteristics of off-Target Sites Bound by the Cas9 Endonuclease." *Nature Biotechnology* 32 (7): 677–83.
- Kuzmichev, Andrei, Raphael Margueron, Alejandro Vaquero, Tanja S. Preissner, Michael Scher, Antonis Kirmizis, Xuesong Ouyang, et al. 2005. "Composition and Histone Substrates of Polycomb Repressive Group Complexes Change during Cellular Differentiation." *Proceedings of the National Academy of Sciences of the United States of America* 102 (6): 1859–64.
- Kuzmichev, Andrei, Kenichi Nishioka, Hediye Erdjument-Bromage, Paul Tempst, and Danny Reinberg. 2002. "Histone Methyltransferase Activity Associated with a Human Multiprotein Complex Containing the Enhancer of Zeste Protein." *Genes & Development* 16 (22): 2893–2905.
- Lachner, M., D. O'Carroll, S. Rea, K. Mechtler, and T. Jenuwein. 2001. "Methylation of Histone H3 Lysine 9 Creates a Binding Site for HP1 Proteins." *Nature* 410 (6824): 116–20.
- Lane, Natasha, Wendy Dean, Sylvia Erhardt, Petra Hajkova, Azim Surani, Jörn Walter, and Wolf Reik. 2003. "Resistance of IAPs to Methylation Reprogramming May Provide a Mechanism for Epigenetic Inheritance in the Mouse." *Genesis* 35 (2): 88–93.
- Latham, John A., and Sharon Y. R. Dent. 2007. "Cross-Regulation of Histone Modifications." *Nature Structural & Molecular Biology* 14 (11): 1017–24.
- Laurent, Louise, Eleanor Wong, Guoliang Li, Tien Huynh, Aristotelis Tsirigos, Chin Thing Ong, Hwee Meng Low, et al. 2010. "Dynamic Changes in the Human Methylome during Differentiation." *Genome Research* 20 (3): 320–31.
- Law, Julie A., and Steven E. Jacobsen. 2010. "Establishing, Maintaining and Modifying DNA Methylation Patterns in Plants and Animals." *Nature Reviews. Genetics* 11 (3): 204–20.
- Leeb, Martin, and Anton Wutz. 2007. "Ring1B Is Crucial for the Regulation of Developmental Control Genes and PRC1 Proteins but Not X Inactivation in Embryonic Cells." *The Journal of Cell Biology* 178 (2): 219–29.
- Lee, Bongyong, and Mark T. Muller. 2009. "SUMOylation Enhances DNA Methyltransferase 1 Activity." *Biochemical Journal* 421 (3): 449–61.
- Lee, Gabsang, Eirini P. Papapetrou, Hyesoo Kim, Stuart M. Chambers, Mark J.

- Tomishima, Christopher A. Fasano, Yosif M. Ganat, et al. 2009. "Modelling Pathogenesis and Treatment of Familial Dysautonomia Using Patient-Specific iPSCs." *Nature* 461 (7262): 402–6.
- Lee, Heather J., Timothy A. Hore, and Wolf Reik. 2014. "Reprogramming the Methylome: Erasing Memory and Creating Diversity." *Cell Stem Cell* 14 (6): 710–19.
- Leonhardt, H., A. W. Page, H. U. Weier, and T. H. Bestor. 1992. "A Targeting Sequence Directs DNA Methyltransferase to Sites of DNA Replication in Mammalian Nuclei." *Cell* 71 (5): 865–73.
- Liang, Gangning, Joy C. Y. Lin, Vivian Wei, Christine Yoo, Jonathan C. Cheng, Carvell T. Nguyen, Daniel J. Weisenberger, et al. 2004. "Distinct Localization of Histone H3 Acetylation and H3-K4 Methylation to the Transcription Start Sites in the Human Genome." *Proceedings of the National Academy of Sciences of the United States of America* 101 (19): 7357–62.
- Li, E., T. H. Bestor, and R. Jaenisch. 1992. "Targeted Mutation of the DNA Methyltransferase Gene Results in Embryonic Lethality." *Cell* 69 (6): 915–26.
- Li, Huipeng, Elise T. Courtois, Debarka Sengupta, Yuliana Tan, Kok Hao Chen, Jolene Jie Lin Goh, Say Li Kong, et al. 2017. "Reference Component Analysis of Single-Cell Transcriptomes Elucidates Cellular Heterogeneity in Human Colorectal Tumors." *Nature Genetics*, March. doi:10.1038/ng.3818.
- Liu, Xiaoli, Qinqin Gao, Pishun Li, Qian Zhao, Jiqin Zhang, Jiwen Li, Haruhiko Koseki, and Jiemin Wong. 2013. "UHRF1 Targets DNMT1 for DNA Methylation through Cooperative Binding of Hemi-Methylated DNA and Methylated H3K9." *Nature Communications* 4: 1563.
- Liu, Yidan, Bin Zhang, Henry Kuang, Gautam Korakavi, Lin-Yu Lu, and Xiaochun Yu. 2016. "Zinc Finger Protein 618 Regulates the Function of UHRF2 (Ubiquitin-like with PHD and Ring Finger Domains 2) as a Specific 5-Hydroxymethylcytosine Reader." *The Journal of Biological Chemistry* 291 (26): 13679–88.
- Li, Xiajun, Mitsuteru Ito, Fen Zhou, Neil Youngson, Xiaopan Zuo, Philip Leder, and Anne C. Ferguson-Smith. 2008. "A Maternal-Zygotic Effect Gene, Zfp57, Maintains Both Maternal and Paternal Imprints." *Developmental Cell* 15 (4): 547–57.
- Li, Xiang, Wei Wei, Qiong-Yi Zhao, Jocelyn Widagdo, Danay Baker-Andresen, Charlotte R. Flavell, Ana D'Alessio, Yi Zhang, and Timothy W. Bredy. 2014. "Neocortical Tet3-Mediated Accumulation of 5-Hydroxymethylcytosine Promotes Rapid Behavioral Adaptation." *Proceedings of the National Academy of Sciences of the United States of America* 111 (19): 7120–25.
- Lorsbach, R. B., J. Moore, S. Mathew, S. C. Raimondi, S. T. Mukatira, and J. R. Downing. 2003. "TET1, a Member of a Novel Protein Family, Is Fused to MLL in Acute Myeloid Leukemia Containing the t(10;11)(q22;q23)." *Leukemia* 17 (3): 637–41.
- Luger, K., T. J. Rechsteiner, A. J. Flaus, M. M. Waye, and T. J. Richmond. 1997. "Characterization of Nucleosome Core Particles Containing Histone Proteins Made in Bacteria." *Journal of Molecular Biology* 272 (3): 301–11.
- Luger, K., and T. J. Richmond. 1998. "The Histone Tails of the Nucleosome." *Current Opinion in Genetics & Development* 8 (2): 140–46.
- Macaulay, Iain C., and Thierry Voet. 2014. "Single Cell Genomics: Advances and Future Perspectives." *PLoS Genetics* 10 (1): e1004126.
- Ma, Dengke K., Maria Carolina Marchetto, Junjie U. Guo, Guo-Li Ming, Fred H. Gage, and Hongjun Song. 2010. "Epigenetic Choreographers of Neurogenesis in the Adult Mammalian Brain." *Nature Neuroscience* 13 (11): 1338–44.
- Maiti, Atanu, and Alexander C. Drohat. 2011. "Thymine DNA Glycosylase Can Rapidly Excise 5-Formylcytosine and 5-Carboxylcytosine: Potential Implications for Active Demethylation of CpG Sites." *The Journal of Biological Chemistry* 286 (41): 35334–38.
- Mali, Prashant, Kevin M. Esvelt, and George M. Church. 2013. "Cas9 as a Versatile Tool for Engineering Biology." *Nature Methods* 10 (10): 957–63.
- Mantel, Charlie, and Hal E. Broxmeyer. 2008. "Sirtuin 1, Stem Cells, Aging, and

- Stem Cell Aging." *Current Opinion in Hematology* 15 (4): 326–31.
- Margot, J. B., A. M. Aguirre-Arteta, B. V. Di Giacco, S. Pradhan, R. J. Roberts, M. C. Cardoso, and H. Leonhardt. 2000. "Structure and Function of the Mouse DNA Methyltransferase Gene: Dnmt1 Shows a Tripartite Structure." *Journal of Molecular Biology* 297 (2): 293–300.
- Marinov, Georgi K., Brian A. Williams, Ken McCue, Gary P. Schroth, Jason Gertz, Richard M. Myers, and Barbara J. Wold. 2014. "From Single-Cell to Cell-Pool Transcriptomes: Stochasticity in Gene Expression and RNA Splicing." *Genome Research* 24 (3): 496–510.
- Martins-Taylor, Kristen, Diane I. Schroeder, Janine M. LaSalle, Marc Lalande, and Ren-He Xu. 2012. "Role of DNMT3B in the Regulation of Early Neural and Neural Crest Specifiers." *Epigenetics: Official Journal of the DNA Methylation Society* 7 (1): 71–82.
- Maunakea, Alike K., Iouri Chepelev, Kairong Cui, and Keji Zhao. 2013. "Intragenic DNA Methylation Modulates Alternative Splicing by Recruiting MeCP2 to Promote Exon Recognition." *Cell Research* 23 (11): 1256–69.
- Mayer, W., A. Niveleau, J. Walter, R. Fundele, and T. Haaf. 2000. "Demethylation of the Zygotic Paternal Genome." *Nature* 403 (6769): 501–2.
- Meilinger, Daniela, Karin Fellingner, Sebastian Bultmann, Ulrich Rothbauer, Ian Marc Bonapace, Wolfgang E. F. Klinkert, Fabio Spada, and Heinrich Leonhardt. 2009. "Np95 Interacts with de Novo DNA Methyltransferases, Dnmt3a and Dnmt3b, and Mediates Epigenetic Silencing of the Viral CMV Promoter in Embryonic Stem Cells." *EMBO Reports* 10 (11). EMBO Press: 1259–64.
- Miller, Jeffrey C., Michael C. Holmes, Jianbin Wang, Dmitry Y. Guschin, Ya-Li Lee, Igor Rupniewski, Christian M. Beausejour, et al. 2007. "An Improved Zinc-Finger Nuclease Architecture for Highly Specific Genome Editing." *Nature Biotechnology* 25 (7): 778–85.
- Moghadam, Keivan Kaveh, Fabio Pizza, Chiara La Morgia, Christian Franceschini, Caterina Tonon, Raffaele Lodi, Piero Barboni, et al. 2014. "Narcolepsy Is a Common Phenotype in HSAN IE and ADCA-DN." *Brain: A Journal of Neurology* 137 (6). Oxford Univ Press: 1643–55.
- Monk, M., M. Boubelik, and S. Lehnert. 1987. "Temporal and Regional Changes in DNA Methylation in the Embryonic, Extraembryonic and Germ Cell Lineages during Mouse Embryo Development." *Development* 99 (3): 371–82.
- Mora-Castilla, Sergio, Cuong To, Soheila Vaezeslami, Robert Morey, Srimeenakshi Srinivasan, Jennifer N. Dumdie, Heidi Cook-Andersen, Joby Jenkins, and Louise C. Laurent. 2016. "Miniaturization Technologies for Efficient Single-Cell Library Preparation for Next-Generation Sequencing." *Journal of Laboratory Automation* 21 (4): 557–67.
- Morey, Lluís, and Kristian Helin. 2010. "Polycomb Group Protein-Mediated Repression of Transcription." *Trends in Biochemical Sciences* 35 (6): 323–32.
- Mori, Tsutomu, Daisuke D. Ikeda, Toshihiko Fukushima, Seiichi Takenoshita, and Hideo Kochi. 2011. "NIRF Constitutes a Nodal Point in the Cell Cycle Network and Is a Candidate Tumor Suppressor." *Cell Cycle* 10 (19): 3284–99.
- Mori, Tsutomu, Daisuke D. Ikeda, Yoshiki Yamaguchi, Motoko Unoki, and NIRF Project. 2012. "NIRF/UHRF2 Occupies a Central Position in the Cell Cycle Network and Allows Coupling with the Epigenetic Landscape." *FEBS Letters* 586 (11): 1570–83.
- Mulholland, Christopher B., Martha Smets, Elisabeth Schmidtman, Susanne Leidescher, Yolanda Markaki, Mario Hofweber, Weihua Qin, et al. 2015. "A Modular Open Platform for Systematic Functional Studies under Physiological Conditions." *Nucleic Acids Research* 43 (17): e112.
- Muto, M. 2002. "Targeted Disruption of Np95 Gene Renders Murine Embryonic Stem Cells Hypersensitive to DNA Damaging Agents and DNA Replication Blocks." *The Journal of Biological Chemistry* 277 (37): 34549–55.
- Myers, F. A., D. R. Evans, A. L. Clayton, A. W. Thorne, and C. Crane-Robinson. 2001. "Targeted and Extended Acetylation of Histones H4 and H3 at Active and Inactive Genes in Chicken Embryo Erythrocytes." *The Journal of Biological*

- Chemistry* 276 (23): 20197–205.
- Neri, Francesco, Danny Incarnato, Anna Krepelova, Stefania Rapelli, Francesca Anselmi, Caterina Parlato, Claudio Medana, Federica Dal Bello, and Salvatore Oliviero. 2015. “Single-Base Resolution Analysis of 5-Formyl and 5-Carboxyl Cytosine Reveals Promoter DNA Methylation Dynamics.” *Cell Reports*, February. doi:10.1016/j.celrep.2015.01.008.
- Neri, Francesco, Anna Krepelova, Danny Incarnato, Mara Maldotti, Caterina Parlato, Federico Galvagni, Filomena Matarese, Hendrik G. Stunnenberg, and Salvatore Oliviero. 2013. “Dnmt3L Antagonizes DNA Methylation at Bivalent Promoters and Favors DNA Methylation at Gene Bodies in ESCs.” *Cell* 155 (1): 121–34.
- Neri, Francesco, Stefania Rapelli, Anna Krepelova, Danny Incarnato, Caterina Parlato, Giulia Basile, Mara Maldotti, Francesca Anselmi, and Salvatore Oliviero. 2017. “Intragenic DNA Methylation Prevents Spurious Transcription Initiation.” *Nature*, February. doi:10.1038/nature21373.
- Ng, H. H., D. N. Ciccone, K. B. Morshead, M. A. Oettinger, and K. Struhl. 2003. “Lysine-79 of Histone H3 Is Hypomethylated at Silenced Loci in Yeast and Mammalian Cells: A Potential Mechanism for Position-Effect Variegation.” *Proceedings of the National Academy of Sciences* 100 (4): 1820–25.
- Ng, H. H., S. Dole, and K. Struhl. 2003. “The Rtf1 Component of the Paf1 Transcriptional Elongation Complex Is Required for Ubiquitination of Histone H2B.” *The Journal of Biological Chemistry* 278 (36): 33625–28.
- Nguyen, Suzanne, Konstantinos Meletis, Dongdong Fu, Sonal Jhaveri, and Rudolf Jaenisch. 2007. “Ablation of de Novo DNA Methyltransferase Dnmt3a in the Nervous System Leads to Neuromuscular Defects and Shortened Lifespan.” *Developmental Dynamics: An Official Publication of the American Association of Anatomists* 236 (6): 1663–76.
- Nishiyama, Atsuya, Luna Yamaguchi, Jafar Sharif, Yoshikazu Johmura, Takeshi Kawamura, Keiko Nakanishi, Shintaro Shimamura, et al. 2013. “Uhrf1-Dependent H3K23 Ubiquitylation Couples Maintenance DNA Methylation and Replication.” *Nature* 502 (7470): 249–53.
- Okano, M., D. W. Bell, D. A. Haber, and E. Li. 1999. “DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for de Novo Methylation and Mammalian Development.” *Cell* 99 (3): 247–57.
- Olek, A., and J. Walter. 1997. “The Pre-Implantation Ontogeny of the H19 Methylation Imprint.” *Nature Genetics* 17 (3): 275–76.
- Papait, Roberto, Christian Pistore, Diego Negri, Daniela Pecoraro, Lisa Cantarini, and Ian Marc Bonapace. 2007. “Np95 Is Implicated in Pericentromeric Heterochromatin Replication and in Major Satellite Silencing.” *Molecular Biology of the Cell* 18 (3): 1098–1106.
- Parekh, Swati, Christoph Ziegenhain, Beate Vieth, Wolfgang Enard, and Ines Hellmann. 2016. “The Impact of Amplification on Differential Expression Analyses by RNA-Seq.” *Scientific Reports* 6 (May): 25533.
- Patil, Vibha, Robyn L. Ward, and Luke B. Hesson. 2014. “The Evidence for Functional Non-CpG Methylation in Mammalian Cells.” *Epigenetics: Official Journal of the DNA Methylation Society* 9 (6): 823–28.
- Pedroso, José Luiz, Orlando Graziani Povoas Barsottini, Ling Lin, Atle Melberg, Acary S. B. Oliveira, and Emmanuel Mignot. 2013. “A Novel de Novo Exon 21 DNMT1 Mutation Causes Cerebellar Ataxia, Deafness, and Narcolepsy in a Brazilian Patient.” *Sleep* 36 (8): 1257–59, 1259A.
- Peng, Lirong, Zhigang Yuan, Hongbo Ling, Kenji Fukasawa, Keith Robertson, Nancy Olashaw, John Koomen, Jiandong Chen, William S. Lane, and Edward Seto. 2011. “SIRT1 Deacetylates the DNA Methyltransferase 1 (DNMT1) Protein and Alters Its Activities.” *Molecular and Cellular Biology* 31 (23): 4720–34.
- Picelli, Simone, Åsa K. Björklund, Omid R. Faridani, Sven Sagasser, Gösta Winberg, and Rickard Sandberg. 2013. “Smart-seq2 for Sensitive Full-Length Transcriptome Profiling in Single Cells.” *Nature Methods* 10 (11): 1096–98.
- Pichler, Garwin, Patricia Wolf, Christine S. Schmidt, Daniela Meilinger, Katrin Schneider, Carina Frauer, Karin Fellinger, Andrea Rottach, and Heinrich

- Leonhardt. 2011. "Cooperative DNA and Histone Binding by Uhrf2 Links the Two Major Repressive Epigenetic Pathways." *Journal of Cellular Biochemistry* 112 (9): 2585–93.
- Pickart, Cecile M. 2004. "Back to the Future with Ubiquitin." *Cell* 116 (2): 181–90.
- Pickart, Cecile M., and David Fushman. 2004. "Polyubiquitin Chains: Polymeric Protein Signals." *Current Opinion in Chemical Biology* 8 (6): 610–16.
- Pogribny, Igor P., and Frederick A. Beland. 2009. "DNA Hypomethylation in the Origin and Pathogenesis of Human Diseases." *Cellular and Molecular Life Sciences: CMLS* 66 (14): 2249–61.
- Pokholok, Dmitry K., Christopher T. Harbison, Stuart Levine, Megan Cole, Nancy M. Hannett, Tong Ihn Lee, George W. Bell, et al. 2005. "Genome-Wide Map of Nucleosome Acetylation and Methylation in Yeast." *Cell* 122 (4): 517–27.
- Ponting, Chris P., Peter L. Oliver, and Wolf Reik. 2009. "Evolution and Functions of Long Noncoding RNAs." *Cell* 136 (4): 629–41.
- Povlsen, Lou K., Petra Beli, Sebastian A. Wagner, Sara L. Poulsen, Kathrine B. Sylvestersen, Jon W. Poulsen, Michael L. Nielsen, Simon Bekker-Jensen, Niels Mailand, and Chunaram Choudhary. 2012. "Systems-Wide Analysis of Ubiquitylation Dynamics Reveals a Key Role for PAF15 Ubiquitylation in DNA-Damage Bypass." *Nature Cell Biology* 14 (10): 1089–98.
- Pradhan, Mihika, Pierre-Olivier Estève, Hang Gyeong Chin, Mala Samaranayake, Gun-Do Kim, and Sriharsa Pradhan. 2008. "CXXC Domain of Human DNMT1 Is Essential for Enzymatic Activity." *Biochemistry* 47 (38): 10000–9.
- Qin, Weihua, Heinrich Leonhardt, and Garwin Pichler. 2011. "Regulation of DNA Methyltransferase 1 by Interactions and Modifications." *Nucleus* 2 (5): 392–402.
- Qin, Weihua, Heinrich Leonhardt, and Fabio Spada. 2011. "Usp7 and Uhrf1 Control Ubiquitination and Stability of the Maintenance DNA Methyltransferase Dnmt1." *Journal of Cellular Biochemistry* 112 (2): 439–44.
- Qin, Weihua, Patricia Wolf, Nan Liu, Stephanie Link, Martha Smets, Federica La Mastra, Ignasi Forné, et al. 2015. "DNA Methylation Requires a DNMT1 Ubiquitin Interacting Motif (UIM) and Histone Ubiquitination." *Cell Research* 25 (8): 911–29.
- Quenneville, Simon, Gaetano Verde, Andrea Corsinotti, Adamandia Kapopoulou, Johan Jakobsson, Sandra Offner, Ilaria Baglivo, et al. 2011. "In Embryonic Stem Cells, ZFP57/KAP1 Recognize a Methylated Hexanucleotide to Affect Chromatin and DNA Methylation of Imprinting Control Regions." *Molecular Cell* 44 (3): 361–72.
- Rajakumara, Eerappa, Zhentian Wang, Honghui Ma, Lulu Hu, Hao Chen, Yan Lin, Rui Guo, et al. 2011. "PHD Finger Recognition of Unmodified Histone H3R2 Links UHRF1 to Regulation of Euchromatic Gene Expression." *Molecular Cell* 43 (2): 275–84.
- Ramsköld, Daniel, Shujun Luo, Yu-Chieh Wang, Robin Li, Qiaolin Deng, Omid R. Faridani, Gregory A. Daniels, et al. 2012. "Full-Length mRNA-Seq from Single-Cell Levels of RNA and Individual Circulating Tumor Cells." *Nature Biotechnology* 30 (8): 777–82.
- Reik, Wolf. 2007. "Stability and Flexibility of Epigenetic Gene Regulation in Mammalian Development." *Nature* 447 (7143): 425–32.
- Ren, Xiaojun, Claudius Vincenz, and Tom K. Kerppola. 2008. "Changes in the Distributions and Dynamics of Polycomb Repressive Complexes during Embryonic Stem Cell Differentiation." *Molecular and Cellular Biology* 28 (9): 2884–95.
- Robertson, Keith D. 2005. "DNA Methylation and Human Disease." *Nature Reviews. Genetics* 6 (8): 597–610.
- Rothbart, Scott B., Bradley M. Dickson, Michelle S. Ong, Krzysztof Krajewski, Scott Houliston, Dmitri B. Kireev, Cheryl H. Arrowsmith, and Brian D. Strahl. 2013. "Multivalent Histone Engagement by the Linked Tandem Tudor and PHD Domains of UHRF1 Is Required for the Epigenetic Inheritance of DNA Methylation." *Genes & Development* 27 (11): 1288–98.

- Rothbart, Scott B., Krzysztof Krajewski, Nataliya Nady, Wolfram Tempel, Sheng Xue, Aimee I. Badeaux, Dalia Barsyte-Lovejoy, et al. 2012. "Association of UHRF1 with Methylated H3K9 Directs the Maintenance of DNA Methylation." *Nature Structural & Molecular Biology* 19 (11): 1155–60.
- Rottach, Andrea, Carina Frauer, Garwin Pichler, Ian Marc Bonapace, Fabio Spada, and Heinrich Leonhardt. 2010. "The Multi-Domain Protein Np95 Connects DNA Methylation and Histone Modification." *Nucleic Acids Research* 38 (6). Oxford Univ Press: 1796–1804.
- Rottach, Andrea, Heinrich Leonhardt, and Fabio Spada. 2009. "DNA Methylation-Mediated Epigenetic Control." *Journal of Cellular Biochemistry* 108 (1): 43–51.
- Rotthier, Annelies, Jonathan Baets, Vincent Timmerman, and Katrien Janssens. 2012. "Mechanisms of Disease in Hereditary Sensory and Autonomic Neuropathies." *Nature Reviews. Neurology* 8 (2): 73–85.
- Rougier, N., D. Bourc'his, D. M. Gomes, A. Niveleau, M. Plachot, A. Paldi, and E. Viegas-Péquignot. 1998. "Chromosome Methylation Patterns during Mammalian Preimplantation Development." *Genes & Development* 12 (14): 2108–13.
- Rountree, M. R., K. E. Bachman, and S. B. Baylin. 2000. "DNMT1 Binds HDAC2 and a New Co-Repressor, DMAP1, to Form a Complex at Replication Foci." *Nature Genetics* 25 (3): 269–77.
- Roux, Kyle J., Dae In Kim, Manfred Raida, and Brian Burke. 2012. "A Promiscuous Biotin Ligase Fusion Protein Identifies Proximal and Interacting Proteins in Mammalian Cells." *The Journal of Cell Biology* 196 (6): 801–10.
- Russell, John, David Chang, Anna Tretiakova, and Malla Padidam. 2006. "Phage Bxb1 Integrase Mediates Highly Efficient Site-Specific Recombination in Mammalian Cells." *BioTechniques* 40 (4): 460–64.
- Sandberg, Rickard. 2014. "Entering the Era of Single-Cell Transcriptomics in Biology and Medicine." *Nature Methods* 11 (1): 22–24.
- Santos, Fátima, Julian Peat, Heather Burgess, Cristina Rada, Wolf Reik, and Wendy Dean. 2013. "Active Demethylation in Mouse Zygotes Involves Cytosine Deamination and Base Excision Repair." *Epigenetics & Chromatin* 6 (1): 39.
- Santos-Reboucas, C. B., and Mmg Pimentel. 2007. "Implication of Abnormal Epigenetic Patterns for Human Diseases." *European Journal of Human Genetics: EJHG* 15 (1). Nature Publishing Group: 10–17.
- Schermelleh, Lothar, Andrea Haemmer, Fabio Spada, Nicole Rösing, Daniela Meilinger, Ulrich Rothbauer, M. Cristina Cardoso, and Heinrich Leonhardt. 2007. "Dynamics of Dnmt1 Interaction with the Replication Machinery and Its Role in Postreplicative Maintenance of DNA Methylation." *Nucleic Acids Research* 35 (13): 4301–12.
- Schermelleh, Lothar, Fabio Spada, Hariharan P. Easwaran, Kourosh Zolghadr, Jean B. Margot, M. Cristina Cardoso, and Heinrich Leonhardt. 2005. "Trapped in Action: Direct Visualization of DNA Methyltransferase Activity in Living Cells." *Nature Methods* 2 (10): 751–56.
- Schoeftner, Stefan, Schoeftner Stefan, Aditya K. Sengupta, Kubicek Stefan, Mechtler Karl, Spahn Laura, Koseki Haruhiko, Jenuwein Thomas, and Wutz Anton. 2006. "Recruitment of PRC1 Function at the Initiation of X Inactivation Independent of PRC2 and Silencing." *The EMBO Journal* 25 (13): 3110–22.
- Shalem, Ophir, Neville E. Sanjana, Ella Hartenian, Xi Shi, David A. Scott, Tarjei S. Mikkelsen, Dirk Heckl, et al. 2014. "Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells." *Science* 343 (6166): 84–87.
- Sharif, Jafar, Masahiro Muto, Shin-Ichiro Takebayashi, Isao Suetake, Akihiro Iwamatsu, Takaho A. Endo, Jun Shinga, et al. 2007. "The SRA Protein Np95 Mediates Epigenetic Inheritance by Recruiting Dnmt1 to Methylated DNA." *Nature* 450 (7171): 908–12.
- Sharma, Shikhar, Daniel D. De Carvalho, Shinwu Jeong, Peter A. Jones, and Gangning Liang. 2011. "Nucleosomes Containing Methylated DNA Stabilize DNA Methyltransferases 3A/3B and Ensure Faithful Epigenetic Inheritance." *PLoS Genetics* 7 (2): e1001286.
- Smallwood, Sébastien A., Shin-Ichi Tomizawa, Felix Krueger, Nico Ruf, Natasha

- Carli, Anne Segonds-Pichon, Shun Sato, Kenichiro Hata, Simon R. Andrews, and Gavin Kelsey. 2011. "Dynamic CpG Island Methylation Landscape in Oocytes and Preimplantation Embryos." *Nature Genetics* 43 (8): 811–14.
- Smets, Martha, Stephanie Link, Patricia Wolf, Katrin Schneider, Veronica Solis, Joel Ryan, Daniela Meilinger, Weihua Qin, and Heinrich Leonhardt. 2017. "DNMT1 Mutations Found in HSNIE Patients Affect Interaction with UHRF1 and Neuronal Differentiation." *Human Molecular Genetics*, March. doi:10.1093/hmg/ddx057.
- Smith, Zachary D., Michelle M. Chan, Tarjei S. Mikkelsen, Hongcang Gu, Andreas Gnirke, Aviv Regev, and Alexander Meissner. 2012. "A Unique Regulatory Phase of DNA Methylation in the Early Mammalian Embryo." *Nature* 484 (7394): 339–44.
- Smith, Zachary D., and Alexander Meissner. 2013. "DNA Methylation: Roles in Mammalian Development." *Nature Reviews. Genetics* 14 (3): 204–20.
- Song, Chun-Xiao, Keith E. Szulwach, Ye Fu, Qing Dai, Chengqi Yi, Xuekun Li, Yujing Li, et al. 2011. "Selective Chemical Labeling Reveals the Genome-Wide Distribution of 5-Hydroxymethylcytosine." *Nature Biotechnology* 29 (1): 68–72.
- Song, Jikui, Olga Rechko, Timothy H. Bestor, and Dinshaw J. Patel. 2011. "Structure of DNMT1-DNA Complex Reveals a Role for Autoinhibition in Maintenance DNA Methylation." *Science* 331 (6020): 1036–40.
- Spada, Fabio, Andrea Haemmer, David Kuch, Ulrich Rothbauer, Lothar Schermelleh, Elisabeth Kremmer, Thomas Carell, Gernot Längst, and Heinrich Leonhardt. 2007. "DNMT1 but Not Its Interaction with the Replication Machinery Is Required for Maintenance of DNA Methylation in Human Cells." *The Journal of Cell Biology* 176 (5): 565–71.
- Spruijt, Cornelia G., Felix Gnerlich, Arne H. Smits, Toni Pfaffeneder, Pascal W. T. C. Jansen, Christina Bauer, Martin Münzel, et al. 2013. "Dynamic Readers for 5-(hydroxy)methylcytosine and Its Oxidized Derivatives." *Cell* 152 (5): 1146–59.
- Steensel, B. van, J. Delrow, and S. Henikoff. 2001. "Chromatin Profiling Using Targeted DNA Adenine Methyltransferase." *Nature Genetics* 27 (3): 304–8.
- Strahl, B. D., and C. D. Allis. 2000. "The Language of Covalent Histone Modifications." *Nature* 403 (6765): 41–45.
- Sun, Lijun, and Zhijian J. Chen. 2004. "The Novel Functions of Ubiquitination in Signaling [Curr. Opin. Cell Biol. 16 (2004) 119]." *Current Opinion in Cell Biology* 16 (3): 339–40.
- Sun, Zhifu, Yanhong Wu, Tamas Ordog, Saurabh Baheti, Jinfu Nie, Xiaohui Duan, Kaori Hojo, Jean-Pierre Kocher, Peter J. Dyck, and Christopher J. Klein. 2014. "Aberrant Signature Methylome by DNMT1 Hot Spot Mutation in Hereditary Sensory and Autonomic Neuropathy 1E." *Epigenetics: Official Journal of the DNA Methylation Society* 9 (8): 1184–93.
- Syeda, Farisa, Rebecca L. Fagan, Matthew Wean, George V. Avvakumov, John R. Walker, Sheng Xue, Sirano Dhe-Paganon, and Charles Brenner. 2011. "The Replication Focus Targeting Sequence (RFTS) Domain Is a DNA-Competitive Inhibitor of Dnmt1." *The Journal of Biological Chemistry* 286 (17): 15344–51.
- Szulwach, Keith E., Xuekun Li, Yujing Li, Chun-Xiao Song, Hao Wu, Qing Dai, Hasan Irier, et al. 2011. "5-hmC-Mediated Epigenetic Dynamics during Postnatal Neurodevelopment and Aging." *Nature Neuroscience* 14 (12): 1607–16.
- Szwagierczak, Aleksandra, Sebastian Bultmann, Christine S. Schmidt, Fabio Spada, and Heinrich Leonhardt. 2010. "Sensitive Enzymatic Quantification of 5-Hydroxymethylcytosine in Genomic DNA." *Nucleic Acids Research* 38 (19): e181.
- Tachibana, Makoto, Yasuko Matsumura, Mikiko Fukuda, Hiroshi Kimura, and Yoichi Shinkai. 2008. "G9a/GLP Complexes Independently Mediate H3K9 and DNA Methylation to Silence Transcription." *The EMBO Journal* 27 (20): 2681–90.
- Tahiliani, Mamta, Kian Peng Koh, Yinghua Shen, William A. Pastor, Hozefa Bandukwala, Yevgeny Brudno, Suneet Agarwal, et al. 2009. "Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1." *Science* 324 (5929): 930–35.

- Takeshita, Kohei, Isao Suetake, Eiki Yamashita, Michihiro Suga, Hirotaka Narita, Atsushi Nakagawa, and Shoji Tajima. 2011. "Structural Insight into Maintenance Methylation by Mouse DNA Methyltransferase 1 (Dnmt1)." *Proceedings of the National Academy of Sciences of the United States of America* 108 (22): 9055–59.
- Tatton-Brown, Katrina, Sheila Seal, Elise Ruark, Jenny Harmer, Emma Ramsay, Silvana Del Vecchio Duarte, Anna Zachariou, et al. 2014. "Mutations in the DNA Methyltransferase Gene DNMT3A Cause an Overgrowth Syndrome with Intellectual Disability." *Nature Genetics* 46 (4): 385–88.
- Tavares, Lúcia, Emilia Dimitrova, David Oxley, Judith Webster, Raymond Poot, Jeroen Demmers, Karel Bezstarosti, et al. 2012. "RYBP-PRC1 Complexes Mediate H2A Ubiquitylation at Polycomb Target Sites Independently of PRC2 and H3K27me3." *Cell* 148 (4): 664–78.
- Taverna, Sean D., Haitao Li, Alexander J. Ruthenburg, C. David Allis, and Dinshaw J. Patel. 2007. "How Chromatin-Binding Modules Interpret Histone Modifications: Lessons from Professional Pocket Pickers." *Nature Structural & Molecular Biology* 14 (11): 1025–40.
- Thoma, F., T. Koller, and A. Klug. 1979. "Involvement of Histone H1 in the Organization of the Nucleosome and of the Salt-Dependent Superstructures of Chromatin." *The Journal of Cell Biology* 83 (2 Pt 1): 403–27.
- Tsai, Wen-Wei, Zhanxin Wang, Teresa T. Yiu, Kadir C. Akdemir, Weiya Xia, Stefan Winter, Cheng-Yu Tsai, et al. 2010. "TRIM24 Links a Non-Canonical Histone Signature to Breast Cancer." *Nature* 468 (7326): 927–32.
- Udeshi, Namrata D., Philipp Mertins, Tanya Svinkina, and Steven A. Carr. 2013. "Large-Scale Identification of Ubiquitination Sites by Mass Spectrometry." *Nature Protocols* 8 (10): 1950–60.
- Uemura, T., E. Kubo, Y. Kanari, T. Ikemura, K. Tatsumi, and M. Muto. 2000. "Temporal and Spatial Localization of Novel Nuclear Protein NP95 in Mitotic and Meiotic Cells." *Cell Structure and Function* 25 (3): 149–59.
- Vieth, Beate, Christoph Ziegenhain, Swati Parekh, Wolfgang Enard, and Ines Hellmann. 2017. "powsimR: Power Analysis for Bulk and Single Cell RNA-Seq Experiments." *bioRxiv*. doi:10.1101/117150.
- Waddington, Conrad Hall. 1957. *The Strategy of the Genes, a Discussion of Some Aspects of Theoretical Biology*, by C.H. Waddington, ... With an Appendix [Some Physico-Chemical Aspects of Biological Organisation] by H. Kacser, G. Allen and Unwin.
- Wang, Liangjun, Wang Liangjun, Brown J. Lesley, Cao Ru, Zhang Yi, Judith A. Kassis, and Richard S. Jones. 2004. "Hierarchical Recruitment of Polycomb Group Silencing Complexes." *Molecular Cell* 14 (5): 637–46.
- Wang, Tim, Jenny J. Wei, David M. Sabatini, and Eric S. Lander. 2014. "Genetic Screens in Human Cells Using the CRISPR-Cas9 System." *Science* 343 (6166): 80–84.
- Wang, Zhibin, Chongzhi Zang, Jeffrey A. Rosenfeld, Dustin E. Schones, Artem Barski, Suresh Cuddapah, Kairong Cui, et al. 2008. "Combinatorial Patterns of Histone Acetylations and Methylations in the Human Genome." *Nature Genetics* 40 (7): 897–903.
- Wang, Zhiqin, Beisha Tang, Yuquan He, and Peng Jin. 2016. "DNA Methylation Dynamics in Neurogenesis." *Epigenomics* 8 (3): 401–14.
- Watanabe, Daisuke, Isao Suetake, Takashi Tada, and Shoji Tajima. 2002. "Stage- and Cell-Specific Expression of Dnmt3a and Dnmt3b during Embryogenesis." *Mechanisms of Development* 118 (1-2): 187–90.
- Watanabe, Daisuke, Isao Suetake, Shoji Tajima, and Kazunori Hanaoka. 2004. "Expression of Dnmt3b in Mouse Hematopoietic Progenitor Cells and Spermatogonia at Specific Stages." *Gene Expression Patterns: GEP* 5 (1): 43–49.
- Watanabe, D., K. Uchiyama, and K. Hanaoka. 2006. "Transition of Mouse de Novo Methyltransferases Expression from Dnmt3b to Dnmt3a during Neural Progenitor Cell Development." *Neuroscience* 142 (3): 727–37.

- Wen, Lu, Xianlong Li, Liying Yan, Yuexi Tan, Rong Li, Yangyu Zhao, Yan Wang, et al. 2014. "Whole-Genome Analysis of 5-Hydroxymethylcytosine and 5-Methylcytosine at Base Resolution in the Human Brain." *Genome Biology* 15 (3): R49.
- Winkelman, Juliane, Ling Lin, Barbara Schormair, Birgitte R. Kornum, Juliette Faraco, Giuseppe Plazzi, Atle Melberg, et al. 2012. "Mutations in DNMT1 Cause Autosomal Dominant Cerebellar Ataxia, Deafness and Narcolepsy." *Human Molecular Genetics* 21 (10): 2205–10.
- Wolf, S. F., D. J. Jolly, K. D. Lunnen, T. Friedmann, and B. R. Migeon. 1984. "Methylation of the Hypoxanthine Phosphoribosyltransferase Locus on the Human X Chromosome: Implications for X-Chromosome Inactivation." *Proceedings of the National Academy of Sciences of the United States of America* 81 (9): 2806–10.
- Woodcock, C. L., and R. P. Ghosh. 2010. "Chromatin Higher-Order Structure and Dynamics." *Cold Spring Harbor Perspectives in Biology* 2 (5): a000596–a000596.
- Workman, J. L., and R. E. Kingston. 1998. "Alteration of Nucleosome Structure as a Mechanism of Transcriptional Regulation." *Annual Review of Biochemistry* 67: 545–79.
- Wossidlo, Mark, Toshinobu Nakamura, Konstantin Lepikhov, C. Joana Marques, Valeri Zakhartchenko, Michele Boiani, Julia Arand, Toru Nakano, Wolf Reik, and Jörn Walter. 2011. "5-Hydroxymethylcytosine in the Mammalian Zygote Is Linked with Epigenetic Reprogramming." *Nature Communications* 2: 241.
- Wu, Angela R., Norma F. Neff, Tomer Kalisky, Piero Dalerba, Barbara Treutlein, Michael E. Rothenberg, Francis M. Mburu, et al. 2014. "Quantitative Assessment of Single-Cell RNA-Sequencing Methods." *Nature Methods* 11 (1): 41–46.
- Wu, Hao, Volkan Coskun, Jifang Tao, Wei Xie, Weihong Ge, Kazuaki Yoshikawa, En Li, Yi Zhang, and Yi Eve Sun. 2010. "Dnmt3a-Dependent Nonpromoter DNA Methylation Facilitates Transcription of Neurogenic Genes." *Science* 329 (5990): 444–48.
- Wu, Hao, and Yi Zhang. 2014. "Reversing DNA Methylation: Mechanisms, Genomics, and Biological Functions." *Cell* 156 (1-2): 45–68.
- Wu, Xuebing, David A. Scott, Andrea J. Kriz, Anthony C. Chiu, Patrick D. Hsu, Daniel B. Dadon, Albert W. Cheng, et al. 2014. "Genome-Wide Binding of the CRISPR Endonuclease Cas9 in Mammalian Cells." *Nature Biotechnology* 32 (7): 670–76.
- Wu, Zhourui, Kevin Huang, Juehua Yu, Thuc Le, Masakazu Namihira, Yupeng Liu, Jun Zhang, Zhigang Xue, Liming Cheng, and Guoping Fan. 2012. "Dnmt3a Regulates Both Proliferation and Differentiation of Mouse Neural Stem Cells." *Journal of Neuroscience Research* 90 (10): 1883–91.
- Xie, Chanlu, Mu Yao, and Qihan Dong. 2014. "Proliferating Cell Nuclear Antigen-Associated Factor (PAF15): A Novel Oncogene." *The International Journal of Biochemistry & Cell Biology* 50 (May): 127–31.
- Xu, G. L., T. H. Bestor, D. Bourc'his, C. L. Hsieh, N. Tommerup, M. Bugge, M. Hulten, X. Qu, J. J. Russo, and E. Viegas-Péquignot. 1999. "Chromosome Instability and Immunodeficiency Syndrome Caused by Mutations in a DNA Methyltransferase Gene." *Nature* 402 (6758): 187–91.
- Xu, Zhengyao, Louise Thomas, Ben Davies, Ronald Chalmers, Maggie Smith, and William Brown. 2013. "Accuracy and Efficiency Define Bxb1 Integrase as the Best of Fifteen Candidate Serine Recombinases for the Integration of DNA into the Human Genome." *BMC Biotechnology* 13 (October): 87.
- Yao, Bing, and Peng Jin. 2014. "Unlocking Epigenetic Codes in Neurogenesis." *Genes & Development* 28 (12): 1253–71.
- Yoder, J. A., N. S. Soman, G. L. Verdine, and T. H. Bestor. 1997. "DNA (cytosine-5)-Methyltransferases in Mouse Cells and Tissues. Studies with a Mechanism-Based Probe." *Journal of Molecular Biology* 270 (3): 385–95.
- Yuan, Junhui, Yujiro Higuchi, Tatsui Nagado, Satoshi Nozuma, Tomonori Nakamura,

- Eiji Matsuura, Akihiro Hashiguchi, Yusuke Sakiyama, Akiko Yoshimura, and Hiroshi Takashima. 2013. "Novel Mutation in the Replication Focus Targeting Sequence Domain of DNMT1 Causes Hereditary Sensory and Autonomic Neuropathy IE." *Journal of the Peripheral Nervous System: JPNS* 18 (1): 89–93.
- Yu, Nam-Kyung, Sung Hee Baek, and Bong-Kiun Kaang. 2011. "DNA Methylation-Mediated Control of Learning and Memory." *Molecular Brain* 4 (January): 5.
- Zamudio, N., and D. Bourc'his. 2010. "Transposable Elements in the Mammalian Germline: A Comfortable Niche or a Deadly Trap?" *Heredity* 105 (1): 92–104.
- Zee, Barry M., Rebecca S. Levin, Bo Xu, Gary LeRoy, Ned S. Wingreen, and Benjamin A. Garcia. 2010. "In Vivo Residue-Specific Histone Methylation Dynamics." *The Journal of Biological Chemistry* 285 (5): 3341–50.
- Zentner, Gabriel E., and Steven Henikoff. 2013. "Regulation of Nucleosome Dynamics by Histone Modifications." *Nature Structural & Molecular Biology* 20 (3): 259–66.
- Zhang, Jiqin, Qinqin Gao, Pishun Li, Xiaoli Liu, Yuanhui Jia, Weicheng Wu, Jiwen Li, Shuo Dong, Haruhiko Koseki, and Jiemin Wong. 2011. "S Phase-Dependent Interaction with DNMT1 Dictates the Role of UHRF1 but Not UHRF2 in DNA Methylation Maintenance." *Cell Research* 21 (12): 1723–39.
- Zhang, Run-Rui, Qing-Yan Cui, Kiyohito Murai, Yen Ching Lim, Zachary D. Smith, Shengnan Jin, Peng Ye, et al. 2013. "Tet1 Regulates Adult Hippocampal Neurogenesis and Cognition." *Cell Stem Cell* 13 (2): 237–45.
- Zhang, Tianyi, Sarah Cooper, and Neil Brockdorff. 2015. "The Interplay of Histone Modifications - Writers That Read." *EMBO Reports* 16 (11): 1467–81.
- Zhao, Chunmei, Wei Deng, and Fred H. Gage. 2008. "Mechanisms and Functional Implications of Adult Neurogenesis." *Cell* 132 (4): 645–60.
- Zhou, Ting, Jun Xiong, Mingzhu Wang, Na Yang, Jiemin Wong, Bing Zhu, and Rui-Ming Xu. 2014. "Structural Basis for Hydroxymethylcytosine Recognition by the SRA Domain of UHRF2." *Molecular Cell* 54 (5): 879–86.

4.2 Abbreviations

Abbreviation	Meaning
5-aza-dC	5-aza-deoxycidine
5caC	5-carboxycytosine
5fC	5-formylcytosine
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
ac	acetylated
ADCA-DN	autosomal dominant cerebellar ataxia, deafness and narcolepsy
AML	acute myeloid leukemia
BAH1/2	bromo-adjacent homology 1/2
BER	base excision repair
CDK2	cyclin-dependent kinase 2
CGI	CpG island
Ch	cherry
CNS	central nervous system
CTD	carboxy-terminal domain
CXXC	zinc finger
DMAP1	DNA methyltransferase associated protein 1

DNMT1/2/3A/3B/3L	DNA methyltransferase 1/2/3A/3B/3L
DNMT1o	oocyte-specific isoform of DNMT1
DUBs	deubiquitinases
ESC	embryonic stem cell
EZH2	enhancer of zeste homolog 2
F3H	fluorescence three hybrid assay
FACS	fluorescence activated cell sorting
GADD45B	growth arrest and DNA damage-inducible protein 45B
GFP	green fluorescent protein
HAT	histone acetyltransferase
HAUSP	herpes virus associated ubiquitin-specific protease
HDAC	histone deacetylase
HMT	histone methyltransferase
HP1	heterochromatin protein 1
HSAN	hereditary sensory and autonomic neuropathy
IAP	intracisternal A particle
ICBP90	inverted CCAAT binding protein of 90 kDa
ICF	Immunodeficiency, Centromeric region instability, Facial anomalies
iPSC	induced pluripotent stem cell

K	lysine
KG	lysine-glycine
M phase	mitosis phase
me1	monomethylated
me2	dimethylated
me3	trimethylated
MEF	mouse embryonic fibroblast
N-terminal	Amino-terminal
NLS	nuclear localization signal
NPC	neuronal progenitor cell
NSC	neuronal stem cell
NSD1/2/3	nuclear receptor SET domain-containing 1/2/3
NTD	N-terminal domain
Oct4	octamer binding transcription factor 4
PARP1	poly(ADP-ribose) polymerase 1
PARylation	poly(ADP-)ribosylation
PBD	PCNA binding domain
PBR	polybasic region
PCNA	proliferating cell nuclear antigen

PCNP	PEST-containing nuclear protein
PGC	primordial germ cell
PH	pericentromeric heterochromatin
PHD	plant homeodomain
PI5P	phosphatidylinositol 5-phosphate
PNS	peripheral nervous system
PRC	polycomb group repressor complex
PTMs	posttranslational modifications
PWWP	proline-tryptophan-tryptophan-proline
RING	really interesting new gene
S phase	synthesis phase
SAH	S-adenosyl- <i>L</i> -homocystein
SAM	S-adenosyl- <i>L</i> -methionine
SETD1A/1B/3	SET domain containing 1A/1B/3
SRA	SET and RING-associated
SUV39H1/H2	suppressor of variegation 3-9 homolog 1 and 2
SUV4-20H1/2	suppressor of variegation 4-20 homolog 1/2
SUZ12	suppressor of zeste 12 protein homolog
SVZ	subventricular zone

TDG	thymine DNA glycosylase
TET	ten eleven translocation
TIP60	Tat interacting protein of 60 kDa
TKO	triple knockout
TS	targeting sequence
TSS	transcriptional start site
TTD	tandem Tudor domain
ub	ubiquitinated
Ubl	ubiquitin-like
UHRF1	ubiquitin-like, containing PHD and RING finger domains 1
UIM	ubiquitin interacting motif
USP7	ubiquitin-specific processing protease 7
ZNF	Zinc finger protein

4.3 Supplementary Material

Supplementary Table. Overview of mutations associated with HSN1E and ADCA-DN. Point mutations and one deletion mutation found in the DNMT1 TS domain of HSN1E and ADCA-DN patients. Investigated molecular dysfunctions are summarized.

Disease	Mutation	Molecular dysfunction	Reference
HSAN1E	D490E + P491Y	Global hypomethylation, site-specific hypermethylation, DNMT1 : misfolding, premature degradation, reduced enzymatic activity, impaired heterochromatin binding during G2	(Klein et al. 2011, 2013)
	Y495C		
	Y495H		
	H569R	Not investigated yet	(David, John, and Emmanuel 2013; Yuan et al. 2013)
	P507N	Not investigated yet	(Moghadam et al. 2014)
	K521Δ		
	T481P	DNMT1 : Loss of heterochromatin binding ability during G2 phase, imbalanced protein homeostasis, translocation of mutant protein to the cytoplasm, aggresome formation and autophagy	(Baets et al. 2015)
	P491L		
	Y524D		
	I531N		
	C353F		
	Y495C		
	Y495H		
ADCA-DN	A570V	Not investigated yet	(Winkelmann et al. 2012; Moghadam et al. 2014)
	G605A		
	V606F		
	C596R	Not investigated yet	(Pedroso et al. 2013)

4.4 Curriculum vitae

Personal information

Name	Martha Smets
Date of birth	01.10.1986
Place of birth	Bonn
Nationality	German
Languages	German (mother tongue), English (fluent), French (basic knowledge), Russian (basic knowledge)

Education

03/2013 – 12/2017	PhD in Biology, LMU Munich
10/2009 – 12/2012	Master of Science in Biology, LMU Munich
10/2007 – 12/2010	Bachelor of Science in Biology, LMU Munich
06/2007	Allgemeine Hochschulreife

Publications

Zhang, Peng, Anne K. Ludwig, Florian D. Hastert, Cathia Rausch, Anne Lehmkuhl, Ines Hellmann, **Martha Smets**, Heinrich Leonhardt, and M. Cristina Cardoso. 2017. "L1 Retrotransposition Is Activated by Ten-Eleven-Translocation Protein 1 and Repressed by Methyl-CpG Binding Proteins." *Nucleus*, May, 0.

Smets, Martha, Stephanie Link, Patricia Wolf, Katrin Schneider, Veronica Solis, Joel Ryan, Daniela Meilinger, Weihua Qin, and Heinrich Leonhardt. 2017. "DNMT1 Mutations Found in HSNIE Patients Affect Interaction with UHRF1 and Neuronal Differentiation." *Human Molecular Genetics*, March. doi:10.1093/hmg/ddx057.

Ziegenhain, Christoph, Beate Vieth, Swati Parekh, Björn Reinius, Amy Guillaumet-Adkins, **Martha Smets**, Heinrich Leonhardt, Holger Heyn, Ines Hellmann, and Wolfgang Enard. 2017. "Comparative Analysis of Single-Cell RNA Sequencing Methods." *Molecular Cell* 65 (4): 631–43.e4.

Qin, Weihua, Patricia Wolf, Nan Liu, Stephanie Link, **Martha Smets**, Federica La Mastra, Ignasi Forné, et al. 2015. "DNA Methylation Requires a DNMT1 Ubiquitin Interacting Motif (UIM) and Histone Ubiquitination." *Cell Research* 25 (8): 911–29.

Mulholland, Christopher B., **Martha Smets**, Elisabeth Schmidtman, Susanne Leidescher, Yolanda Markaki, Mario Hofweber, Weihua Qin, et al. 2015. "A Modular Open Platform for Systematic Functional Studies under Physiological Conditions." *Nucleic Acids Research* 43 (17): e112.

Gluschnaider, Udi, Rachel Hertz, Sarit Ohayon, Elia Smeir, **Martha Smets**, Eli Pikarsky, and Jacob Bar-Tana. 2014. "Long-Chain Fatty Acid Analogues Suppress Breast Tumorigenesis and Progression." *Cancer Research* 74 (23): 6991–7002.

Manuscripts in preparation:

Martha Smets and Elisabeth Karg, Ignasi Forné, Weihua Qin, Christopher B Mulholland, Joel Ryan, Axel Imhof, Sebastian Bultmann and Heinrich Leonhardt. "Ubiquitome analysis reveals PAF15 as a specific ubiquitination target of UHRF1 in embryonic stem cells." *Manuscript in preparation*.

Anne-Kathrin Garz, Saskia Wolf, Sonja Grath, Verena Gaidzik, Stefan Habringer, Binje Vick, Martina Rudelius, Christoph Ziegenhain, Silvia Herold, Marie-Theresa Weickert, **Martha Smets**, Christian Peschel, Robert Oostendorp, Sebastian Bultmann, Irmela Jeremais, Christian Thiede, Konstanze Döhner, Ulrich Keller, and Katharina Götze. "Azacitidine combined with the selective FLT3 kinase inhibitor crenolanib disrupts stromal protection and inhibits expansion of residual leukemia-initiating cells in FLT3-ITD AML with concurrent epigenetic mutations". *Manuscript submitted*.

Nan Liu, Pin Zhao, **Martha Smets**, Christina Bauer, Christoph Ziegenhain, Sebastian Bultmann, Ignasi Forné, Axel Imhof, Wolfgang Enard, Heinrich Leonhardt and Weihua Qin. "ZNF599 and DNMT3A coordinately control nuclear envelope organization by repression of SUN4 expression". *Manuscript in preparation*.

Weihua Qin, Christoph Mulholland, Nan Liu, Pin Zhao, Sebastian Bultmann, **Martha Smets**, Susanne Leidescher, Christoph Ziegenhain, Wolfgang Enard, Heinrich Leonhardt. "HP1 β phosphorylation promotes exit from naive pluripotency by regulation of glycolytic metabolism." *Manuscript in preparation*.

4.5 Contributions

Declaration of contributions to “DNMT1 mutations found in HSANIE patients affect interaction with UHRF1 and neuronal differentiation”

This project was conceived by Heinrich Leonhardt and Patricia Wolf. Martha Smets performed the experiments shown in the Figures 2C, 3D-F, 6 (in collaboration with Stephanie Link), and Supplementary Figures S4, S6, S7. Stephanie Link performed together with Patricia Wolf experiments for figures: 2A, 3C, S5. Martha Smets wrote the manuscript in together with Stephanie Link and Heinrich Leonhardt. Martha Smets combined all results to design the final Figures, included all changes during the revisions and corrected the manuscript together with Heinrich Leonhardt and Stephanie Link. Moreover, Martha Smets was responsible for submitting and resubmitting the final manuscript as well as corresponding with the editors of the journal “Human Molecular genetics”.

Prof. Dr. Heinrich Leonhardt

Martha Smets

Stephanie Link

Declaration of contributions to “DNA methylation requires a DNMT1 ubiquitin interacting motif (UIM) and histone ubiquitination”

This project was conceived by Heinrich Leonhardt, Weihua Qin and Patricia Wolf. I performed biochemical and cell biological experiments for the following Figures: Supplementary Figures S3B-E (in collaboration with Stephanie Link). I proofread and corrected the manuscript.

Declaration of contributions to “A modular open platform for systematic functional studies under physiological conditions”

This project was conceived by Sebastian Bultmann, Christopher Mulholland and Heinrich Leonhardt. I performed biochemical and cell biological experiments for the following Figures: 1D, 2C, 3F-H. I proofread the manuscript.

Declaration of contributions to “Ubiquitome analysis reveals PAF as a specific ubiquitination target of UHRF1 in embryonic stem cells”

This project was conceived by Elisabeth Karg, Heinrich Leonhardt and me. I wrote the manuscript and designed the Figures together with Elisabeth Karg. I generated and characterized the *UHRF* KO cell lines and performed confocal microscopy experiments. Elisabeth Karg performed proteomics and biochemistry experiments and data analysis. Elisabeth Karg and I proofread and approved the final manuscript.

Prof. Dr. Heinrich Leonhardt

Martha Smets

Elisabeth Karg

Declaration of contributions to “Comparative Analysis of Single-Cell RNA Sequencing Methods”

I performed the cell culture of mESCs for the experiments. Heinrich Leonhardt supervised the experimental work. I proofread and approved the final manuscript.

4.6 Statutory declaration and statement

Declaration according to the 'Promotionsordnung der LMU München für die Fakultät Biologie`.

Betreuung

Hiermit erkläre ich, dass die vorgelegte Arbeit an der LMU von Herrn Prof. Dr. Heinrich Leonhardt betreut wurde.

Anfertigung

Hiermit versichere ich an Eides statt, dass die Dissertation selbstständig und ohne unerlaubte Hilfsmittel angefertigt wurde. Über Beiträge, die im Rahmen der kumulativen Dissertation in Form von Manuskripten in der Dissertation enthalten sind, wurde im Kapitel 4.4 Rechenschaft abgelegt und die eigenen Leistungen wurden aufgelistet.

Prüfung

Hiermit erkläre ich, dass die Dissertation weder als ganzes noch in Teilen an einem anderen Ort einer Prüfungskommission vorgelegt wurde. Weiterhin habe ich weder an einem anderen Ort eine Promotion angestrebt noch angemeldet noch versucht eine Doktorprüfung abzulegen.

München, den 06.06.2017

Martha Smets

